HAL
open science

Compressed sensing with structured sparsity and structured acquisition

Claire Boyer, Jérémie Bigot, Pierre Weiss

To cite this version:

Claire Boyer, Jérémie Bigot, Pierre Weiss. Compressed sensing with structured sparsity and structured acquisition. 2015. hal-01149456v1

HAL Id: hal-01149456

https://hal.science/hal-01149456v1

Preprint submitted on 7 May 2015 (v1), last revised 10 Apr 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Compressed sensing with structured sparsity and structured acquisition

Claire Boyer[(1)], Jérémie Bigot[(2)] and Pierre Weiss[(1,3)]

[(1)] Institut de Mathématiques de Toulouse (UMR 5219), CNRS, Université de Toulouse, France
claire.boyer@math.univ-toulouse.fr
[(2)] Institut de Mathématiques de Bordeaux (UMR 5251), CNRS, Université de Bordeaux, France
jeremie.bigot@math.u-bordeaux1.fr
[(3)] Institut des Technologies Avancées du Vivant (USR 3505), CNRS, Toulouse, France
pierre.armand.weiss@gmail.com

May 7, 2015

Abstract

Compressed Sensing (CS) is an appealing framework for applications such as Magnetic Resonance Imaging (MRI). However, up-to-date, the sensing schemes suggested by CS theories are made of random isolated measurements, which are usually incompatible with the physics of acquisition. To reflect the physical constraints of the imaging device, we introduce the notion of blocks of measurements: the sensing scheme is not a set of isolated measurements anymore, but a set of groups of measurements which may represent any arbitrary shape (radial lines for instance). Structured acquisition with blocks of measurements are easy to implement, and they give good reconstruction results in practice. However, very few results exist on the theoretical guarantees of CS reconstructions in this setting. In this paper, we fill the gap between CS theory and acquisitions made in practice. To this end, the key feature to consider is the structured sparsity of the signal to reconstruct. In this paper, we derive new CS results for structured acquisitions and signal satisfying a prior structured sparsity. The obtained results are RIPless, in the sense that they do not hold for any s-sparse vector, but for sparse vectors with a given support S. Our results are thus support-dependent, and they offer the possibility for flexible assumptions on the structure of S. Moreover, our results are also drawing-dependent, since we highlight an explicit dependency between the probability of reconstructing a sparse vector and the way of choosing the blocks of measurements.

Key-words: Compressed Sensing, blocks of measurements, structured sparsity, MRI, exact recovery, ℓ_1 minimization.

1 Introduction

Since its introduction in [CRT06b, Don06], compressive sampling triggered a massive interest in fundamental and applied research. However, despite recent progresses, existing theories are still insufficient to explain the success of compressed acquisitions in many practical applications. Our aim in this paper is to extend the applicability of the theory by combining two new ingredients: structured sparsity and acquisition structured by blocks.

1.1 A brief review of existing results

Compressed sensing - as proposed in [CRT06a] - consists in recovering a signal $x \in \mathbb{C}^n$, from a vector of measurements $y = Ax$, where $A \in \mathbb{C}^{m \times n}$ is the sensing matrix. Typical theorems

state that if x is s-sparse, $m \gtrsim s\log(n)$ and A have some good features, then x can be recovered exactly from y by solving the following ℓ_1 minimization problem :

$$\min_{x\in\mathbb{C}^n, Ax=y} \|x\|_1. \tag{1}$$

Moreover, it can be shown that the recovery is robust to noise if the constraint in (1) is penalized. An important fact about this theorem is that the number of measurements mostly depends on the intrinsic dimension s rather than the ambient dimension n.

The first sensing matrices studied were generated by selecting a few Fourier coefficients uniformly at random [CRT06a]. The theory was then extended to random matrices with i.i.d. components [CT06] and sensing vectors selected randomly from orthogonal bases [CRT06b] or discrete or continuous frames [CP11]. An interesting class of sensing matrices for application was introduced in [Rom09] and based on a convolution with a random vector. In the meanwhile, different tools were introduced to analyse sensing matrices such as UUP (uniform uncertainty principle), RIP (restricted isometry property). Many recent works on CS are rather based on coherence or local coherence [KW14] and RIPless proofs [CP11]. The book [FR13] proposes a detailed and self-contained description of most of those concepts.

A common aspect of the above results is that they assume no structure - apart from sparsity - in the signals to recover. Recovering arbitrary sparse vectors is a very demanding property that precludes the use of CS in many practical settings. To the best of our knowledge, the work [AHPR13] is the first to consider the recovery of sparse signals with a structured support. To treat such cases, the authors introduce new concepts such as sparsity by levels.

1.2 The need for new results

One of the main current limitations of CS is the small number of sensing matrices studied so far. Let us illustrate this insufficiency with a practical example from Magnetic Resonance Imaging (MRI). This example will be the red thread of the paper.

In MRI, images are sampled in the Fourier domain and can be assumed to be sparse in the wavelet domain. Under this hypothesis, a byproduct of standard compressed sensing results [CP11] implies that variable density sampling [PVW11, CCW13, KW14] allows perfect reconstruction with a limited number of measurements. The theory in [AHPR13], based on structured sparsity, also leads to the same conclusion. This is illustrated in Figure (1). The white dots on the left image indicate which Fourier coefficients are probed. As can be seen, 4.6% of the coefficients are enough to reconstruct a well resolved image.

Unfortunately, probing Fourier coefficients independently at random is infeasible in MRI: the samples have to lie on piecewise smooth trajectories [CCKW14, BWB14]. One of the most successful practical sampling scheme in MRI consists of measuring whole lines of Fourier coefficients at random [LDP07]. The lines are all parallel and drawn indepently, at random, according to a certain distribution. This is illustrated in Figure 2 in 2D [1]. As can be seen on this example, compressed acquisitions with a lot of structure make it possible to reconstruct well resolved images. To the best of our knowledge, there currently exists no theory able to explain this favorable behavior. The only works dealing with such an acquisition are [PDG15, BBW14]. They assume no structure in the sparsity and we showed in [BBW14] that structure was crucially needed to explain results such as those in Figure 2. We will recall this result in Section 4.3.1.

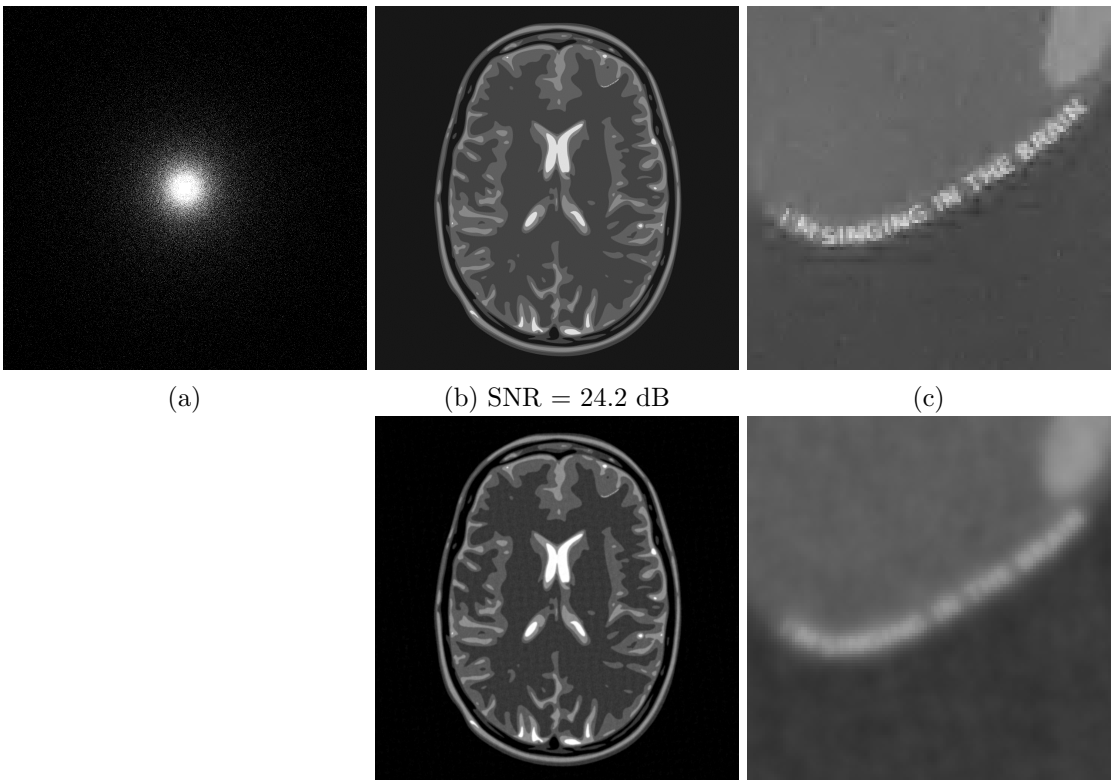
1.3 Contributions

The main contributions of this paper are the following: (i) we provide recovery guarantees for vectors $x \in \mathbb{C}^n$ with a fixed support $S \subset \{1, \ldots, n\}$. This is in strong contrast with the usual works that consider the reconstruction of arbitrary s-sparse vectors. (ii) we provide a theoretical

[1] Paper [LDP07] considers 3D lines.

justification to the use of block acquisitions in compressed sensing. By doing so, we enrich the family of sensing matrices available for compressed sensing. The proposed theory has a few important consequences:

- the concepts of RIP or coherence are not sound anymore. They are replaced by a new quantity $\Gamma(S, \pi)$ which explicitly depends on the support S, the sensing vectors and the drawing probability π of measurements.
- the proposed theory allows envisioning the use of CS in situations that were not possible before. The use of incoherent transforms is not necessary anymore, given that the support S has some good properties.
- The example given in Figure 2 can be analyzed precisely. In particular, we show that a block structured acquisition can be used, only if the support structure is adapted to it. The resulting structures are more complex than the sparsity by levels of [AHPR13].
- The explicit dependency on the support S allows to provide guarantees of reconstruction for random signals with known distribution.



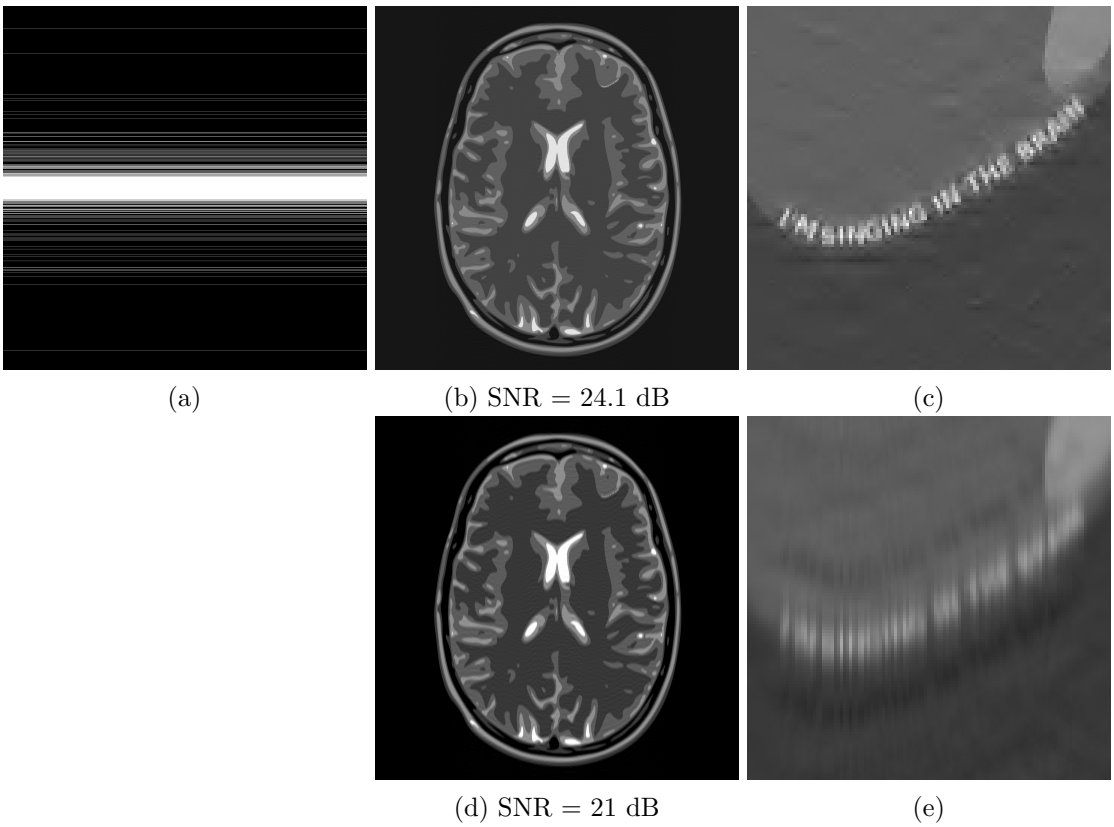
(a) (b) SNR = 24.2 dB (c)

(d) SNR = 21 dB (e)

Figure 1: An example of reconstruction of a 2048×2048 MR image from isolated measurements. (a) Sensing pattern from a variable density sampling strategy (with 4.6% measurements). (b) Corresponding reconstruction via ℓ_1-minimization. (c) A zoom on a part of the reconstructed image. (d) Image obtained by using the pseudo-inverse transform. (e) A zoom on a part of this image.

1.4 Related notions in the literature

In this paper, structured acquisition denotes the constraints imposed by the physics of the acquisition, that are modeled using blocks of measurements extracted from a full deterministic



(a) (b) SNR = 24.1 dB (c)

(d) SNR = 21 dB (e)

Figure 2: An example of reconstruction of a 2048×2048 MR image from blocks of measurements. (a) Sampling pattern horizontal lines (13% of measurements). (b) Corresponding reconstruction via ℓ_1-minimization. (c) A zoom on a part of the reconstructed image. (d) Image obtained by using the pseudo-inverse transform. (e) A zoom on a part of this image.

matrix A_0. This notion of structured acquisition differs from the notion of structured random matrices, as described in [Rau10] and [DE11]. Indeed, this latter strategy is based on acquiring isolated measurements randomly drawn from the rows of a deterministic matrix. The resulting sensing matrix has thus some inherent structure, which is not the case of random matrices with i.i.d. entries, that were initially considered in CS. In our paper, the sensing matrix A is even more structured, in the sense that the full sampling matrix A_0 has been partitioned into blocks of measurements.

We also focus on obtaining RIPless results by combining structured acquisition and structured sparsity. RIPless results [CP11] refer to CS approaches that are non-uniform in the sense they hold for a given sensing matrix A and a given support S of length s, but not for all s-sparse vectors. Nevertheless, existing RIPless results in the literature are only based on the degree of sparsity $s = |S|$. A main novelty of this paper is to develop RIPless results that depend explicitly on the support S (and not only on its length s) of the signal to reconstruct. This strategy allows to incorporate any kind of *prior* information on the structure of S to study its influence on the quality of CS reconstructions. To the best of our knowledge, this setting has not been considered so far. Note that preliminary results have been proposed in [GN08, HSIG13].

In [DE11], a more general model on the signal sparsity is also considered. Indeed, it deals with sparse signals that can be represented in a union of subspaces. Nevertheless, the authors focus on modifying the recovery algorithm to the chosen assumption on sparsity. In this paper, any assumption on the support S of the signal to reconstruct can be addressed, including the case of union of subspaces for instance. Moreover, we do not particularize the reconstruction method. Indeed, instead of modifying the recovery algorithm (i.e. the ℓ_1-minimization problem)

as in [DE11, JKN+12], we focus on adapting the sampling scheme to the assumption made on sparsity. Furthermore, as [AHR14b] suggests, it seems that exploiting structure in sampling is more efficient that exploiting structure in the recovery algorithm.

1.5 Organization of the paper

The paper organization is as follows. Section 2 gives the formal setting of structured acquisition. Section 3 gives the main results, with a precise definition of $\Gamma(S, \pi)$. Applications of our main theorem to various settings are presented in Section 4. Technical appendices contain the proofs of the main results of this paper.

2 Preliminaries

2.1 Notation

In this paper, n denotes the dimension of the signal to reconstruct. The notation $S \subset \{1, \dots, n\}$ refers to the support of the signal to reconstruct. The vectors $(e_i)_{1 \leq i \leq p}$ denote the vectors of the canonical basis of $\mathbb{R}^d$, where d will be equal to n or $\sqrt{n}$, depending on the context. In the sequel, we set $P_S \in \mathbb{R}^{n \times n}$ to be the projection matrix onto span $(\{e_i, i \in S\})$, i.e. the diagonal matrix with the j-th diagonal entry equal to 1 if $j \in S$, and 0 otherwise. We will use the shorthand notation $M_S \in \mathbb{C}^{n \times n}$ and $v_S \in \mathbb{C}^n$ to denote the matrix MP_S and the vector $P_S v$ for $M \in \mathbb{C}^{n \times n}$ and $v \in \mathbb{C}^n$. For any matrix M, for any $1 \leq p, q \leq \infty$, the operator norm $\|M\|_{p \to q}$ is defined as

$$\|M\|_{p \to q} = \sup_{\|v\|_p \leq 1} \|Mv\|_q,$$

with $\|\cdot\|_p$ and $\|\cdot\|_q$ denoting the standard ℓ_p and ℓ_q norms. Note that for a matrix $M \in \mathbb{R}^{n \times n}$,

$$\|M\|_{\infty \to \infty} = \max_{1 \leq i \leq n} \|e_i^* M\|_1.$$

Finally, the function sign : $\mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$(\mathrm{sign}(x))_i = \begin{cases} 1 & \text{if} \quad x_i > 0 \\ -1 & \text{if} \quad x_i < 0 \\ 0 & \text{if} \quad x_i = 0. \end{cases}$$

2.2 Sampling strategy

In this paper, we assume that we are given some orthogonal matrix $A_0 \in \mathbb{C}^{n \times n}$ representing the set of possible linear measurements imposed by a specific sensor device. Let $(\mathcal{I}_k)_{1 \leq k \leq M}$ denote a partition of the set $\{1, \dots, n\}$. The rows $(a_i^*)_{1 \leq i \leq n} \in \mathbb{C}^n$ of A_0 are partitioned into the following blocks dictionary $(B_k)_{1 \leq k \leq M}$, such that

$$B_k = (a_i^*)_{i \in \mathcal{I}_k} \in \mathbb{C}^{|\mathcal{I}_k| \times n} \qquad s.t. \qquad \mathcal{I}_k \subset \{1, \dots, n\},$$

with $\sqcup_{k=1}^M \mathcal{I}_k = \{1, \dots, n\}$. The sensing matrix A is then constructed by randomly drawing blocks as follows

$$A = \frac{1}{\sqrt{m}} \left(\frac{1}{\sqrt{\pi_{K_\ell}}} B_{K_\ell} \right)_{1 \leq \ell \leq m}, \tag{2}$$

where $(K_\ell)_{1 \leq \ell \leq m}$ are i.i.d. copies of a random variable K such that

$$\mathbb{P}(K = k) = \pi_k,$$

for all $1 \leq k \leq M$. Moreover, thanks to the renormalization of the blocks B_{K_ℓ} by the weights $1/\sqrt{\pi_{K_\ell}}$ in model (2), the random block B_K satisfies

$$\mathbb{E}\left(\frac{B_K^* B_K}{\pi_K}\right) = \sum_{k=1}^{M} B_k^* B_k = \mathrm{Id}, \tag{3}$$

since A_0 is orthogonal and $(B_k)_{1\leq k\leq M}$ is a partition of the rows of A_0.

Remark 2.1. *The case of overlapping blocks can also be handled. To do so, we may define the blocks $(B_k)_{1\leq k\leq M}$ as follows:*

$$B_k = \left(\frac{1}{\sqrt{\alpha_i}} a_i^*\right)_{i\in\mathcal{I}_k}, \qquad \textit{for} \quad 1 \leq k \leq M,$$

where $\bigcup_{k=1}^{M} \mathcal{I}_k = \{1, \ldots, n\}$. The coefficients $(\alpha_i)_{1\leq i\leq n}$ denotes the multiplicity of the row a_i^*, namely the number of appearances $\alpha_i = |\{k, i \in \mathcal{I}_k\}|$ of this row in different blocks. This renormalization is sufficient to ensure the isotropy condition $\mathbb{E}\left(\frac{B_K^* B_K}{\pi_K}\right) = \mathrm{Id}$ where K is defined as above.

Note that our block sampling strategy encompasses the standard acquisition based on isolated measurements. Indeed, isolated measurements can be considered as blocks of measurements consisting of only one row of A_0.

Remark 2.2. *More generally, the theorems could be extended - with slight adaptations - to the case where the sensing matrix is*

$$A = \frac{1}{\sqrt{m}} \begin{pmatrix} B_{K_1} \\ \vdots \\ B_{K_m} \end{pmatrix}$$

where $B_{K_1}, \ldots, B_{K_m}$ are i.i.d. copies of a random matrix $B \in \mathbb{C}^{b\times n}$ satisfying

$$\mathbb{E}(B^* B) = \mathrm{Id}.$$

The integer b is itself random and Id *is the $n\times n$ identity matrix. Assuming that B takes its value in a countable family $(B_k)_{k\in\mathcal{K}}$, this formalism covers a large number of applications described in [BBW14]: (i) blocks with i.i.d. entries, (ii) partition of the rows of orthogonal transforms, (iii) cover of the rows of orthogonal transforms, (iv) cover of the rows from tight frames.*

3 Main Results

3.1 Fundamental quantities

Before introducing our main results, we need to define some quantities (reminiscent of the coherence) that will play a key role in our analysis.

Definition 3.1. *Consider a blocks dictionary $(B_k)_{1\leq k\leq M}$. Let $S \subset \{1, \ldots, n\}$ and π be a probability distribution on $\{1, \ldots, M\}$. Define*

$$\Theta(S, \pi) := \max_{1\leq k\leq M} \frac{1}{\pi_k} \|B_k^* B_{k,S}\|_{\infty\to\infty} = \max_{1\leq k\leq M} \max_{1\leq i\leq n} \frac{\|e_i^* B_k^* B_{k,S}\|_1}{\pi_k}, \tag{4}$$

$$\Upsilon(S, \pi) := \max_{1\leq i\leq n} \sup_{\|v\|_\infty\leq 1} \sum_{k=1}^{M} \frac{1}{\pi_k} \left|e_i^* B_k^* B_{k,S} v\right|^2, \tag{5}$$

$$\Gamma(S, \pi) := \max\left(\Upsilon(S, \pi), \Theta(S, \pi)\right). \tag{6}$$

For the sake of readability, we will sometimes use the shorter notation Θ, Υ and Γ to denote $\Theta(S,\pi), \Upsilon(S,\pi)$ and $\Gamma(S,\pi)$. In Definition 3.1, Θ is related to the local coherence and the degree of sparsity, when the blocks are made of only one row (the case of isolated measurements). Indeed, in such a case, Θ reads as follows

$$\Theta(S,\pi) := \max_{1\leq k\leq n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k} \leq s \cdot \max_{1\leq k\leq n} \frac{\|a_k\|_\infty^2}{\pi_k}.$$

The quantity $\max_{1\leq k\leq n} \frac{\|a_k\|_\infty^2}{\pi_k}$ refers to the usual notion of coherence described in [CP11]. The quantity Υ is new and it is more delicate to interpret. It reflects an inter-block coherence. A rough upper-bound for Υ is

$$\Upsilon(S,\pi) \leq \sum_{k=1}^{M} \frac{1}{\pi_k} \|B_k^* B_{k,S}\|_{\infty\to\infty}^2 .$$

by switching the maximum and supremum with the sum in the definition of Υ. However, it is important to keep this order (maximum, supremum and sum) to measure interferences between blocks. In Section 4, we give more precise evaluations of $\Theta(S,\pi)$ and $\Upsilon(S,\pi)$ in particular cases.

Remark 3.2 (Support-dependency and drawing-dependency)**.** *In Definition 3.1, the quantities Θ and Υ are drawing-dependent and support-dependent. Indeed, Γ does not only depend on the degree of sparsity $s = |S|$. To the best of our knowledge, existing theories in CS only rely on s, see [CRT06a, CP11], or on degrees of sparsity structured by levels, see [AHPR13]. Since Γ is explicitly related to S, this allows to incorporate prior assumptions on the structure of S. Besides, the dependency on π (i.e. the way of drawing the measurements) is also explicit in the definition of Γ. This offers the flexibility to analyze the influence of π on the required number of measurements. We therefore believe that the introduced quantities might play an important role in the future analysis of CS.*

3.2 Exact recovery guarantees

Our main result reads as follows.

Theorem 3.3. *Let $S \subset \{1,\dots,n\}$ be a set of indices of cardinality $s \geq 16$ and suppose that $x \in \mathbb{C}^n$ is an s-sparse vector supported on S. Fix $\varepsilon \in (0,1)$. Suppose that the sampling matrix A is constructed as in (2). Suppose that $\Gamma(S,\pi) \geq 1$. If*

$$m \geq 73 \cdot \Gamma(S,\pi) \ln(64s) \left(\ln\left(\frac{9n}{\varepsilon}\right) + \ln\ln(64s) \right), \tag{7}$$

then x is the unique solution of (1) with probability larger than $1-\varepsilon$.

Remark 3.4. *In the sequel, we will simplify condition (7) by writing:*

$$m \geq C \cdot \Gamma(S,\pi) \ln(s) \ln\left(\frac{n}{\varepsilon}\right)$$

where C is a universal constant.

The proof of Theorem 3.3 is contained in AppendixA.1. It relies on the construction of an inexact dual certificate satisfying appropriate properties that are described in Lemma A.1. Then our proof is based on the so-called golfing scheme introduced in [Gro11] for matrix completion, and adapted by [CP11] for compressed sensing from isolated measurements. Nevertheless, the methodology in the proof differs from the techniques that are described in [CP11]. Indeed, a straightforward adaptation of the arguments in [CP11] does not allow to relate the number of measurements to the support S. In the golfing scheme, the main trick is the control of operator

norms of random matrices extracted from the sensing matrix A. In [CP11], it is proposed to control (in probability) the operator norms $\|\cdot\|_{\infty\to 2}$ and $\|\cdot\|_{2\to 2}$. However, this technique only gives results depending on the degree of sparsity s. In order to include an explicit dependency on the support S, one has to modify the golfing scheme in [CP11], by controlling the operator norm $\|\cdot\|_{\infty\to\infty}$, instead of controlling the operator norms $\|\cdot\|_{\infty\to 2}$ and $\|\cdot\|_{2\to 2}$. A similar idea has been developed in [AHPR13] but our main result is more general than the finite-dimensional setting in [AHPR13].

Remark 3.5. *Compared to standard results in compressed sensing, the condition required in Theorem 3.3 involves an extra* $\ln(s)$ *factor. The latter can be removed at the price of additional technicalities, using an extra hypothesis called the Balancing Property in [AHPR13]. In this paper, this possibility has been put aside for the sake of clarity.*

3.3 Consequences for stochastic signal models

The explicit dependency of Γ in S allows us to consider the case of a random support S.

Proposition 3.6. *Let $S \subset \{1,\dots,n\}$ denote a random support. For some real positive γ, suppose that the event $\Gamma(S,\pi) \leq \gamma$ occurs with probability larger than $1-\varepsilon'(\gamma)$. If $m \gtrsim \gamma \ln(s)\ln(n/\varepsilon)$, then x is the unique solution of Problem 1 with probability higher than $1-\varepsilon-\varepsilon\varepsilon'(\gamma)$.*

Proof. Set $m \gtrsim \gamma\ln(s)\ln(n/\varepsilon)$. Define the event R "x is the unique solution of Problem 1" where R stands for "reconstruction of the signal". Define also A the event "$\Gamma(S,\pi) \geq \gamma$". The hypothesis of Proposition 3.6 and Theorem 3.3 give that $\mathbb{P}(R|A) \geq 1-\varepsilon$. To prove Proposition 3.6, we must quantify

$$\begin{aligned}\mathbb{P}(R) &= \mathbb{P}(R\cap A) + \mathbb{P}(R\cap A^c) = \mathbb{P}(R|A)\,\mathbb{P}(A) + \mathbb{P}(R\cap A^c)\\ &\geq (1-\varepsilon)\big(1-\varepsilon'(\gamma)\big) = 1-\varepsilon-\varepsilon\varepsilon'(\gamma),\end{aligned}$$

which concludes the proof. ■

3.4 Choice of the drawing probability

The choice of a drawing probability π minimizing the required number of block measurements in Theorem 3.3, is a delicate issue. The distribution $\pi^\star$ minimizing $\Theta(S,\pi)$ in Equation (4) can be obtained explicitly:

$$\pi_k^\star = \frac{\|B_k^* B_{k,S}\|_{\infty\to\infty}}{\sum_{\ell=1}^M \|B_\ell^* B_{\ell,S}\|_{\infty\to\infty}}, \quad \text{for} \quad 1 \leq k \leq M. \tag{8}$$

Unfortunately, the minimization of $\Upsilon(S,\pi)$ with respect to π seems much more involved and we leave this issue as an open question in the general case.

Note however that in all the examples treated in the paper, we derive upper bounds depending on (S,π) for $\Upsilon(S,\pi)$ and $\Theta(S,\pi)$ that coincide. The distribution $\pi^\star$ is then set to minimize the latter upper bound.

Note also that optimizing π independently of S will result in a sole dependence to the degree of sparsity $s = |S|$ which is not desirable if one wants to exploit structured sparsity.

4 Applications

In this section, we first show that Theorem 3.3 can be used to recover state of the art results in the case of isolated measurements [CP11]. We then show that it allows recovering recent results when a prior on the sparsity structure is available. The proposed setting however applies to a

wider setting even in the case of isolated measurements. Finally, we illustrate the consequences of our results when the acquisition is constrained by blocks of measurements. In the latter case, we show that the sparsity structure should be adapted to the sampling structure for exact recovery.

4.1 Isolated measurements with arbitrary support

First, we focus on an acquisition based on isolated measurements which is the most widespread in CS. This case corresponds to choose blocks of form $B_k = a_k^*$ for $1 \leq k \leq n$ with $M = n$, where a_k^* are the rows of an orthogonal matrix. In such a setting, the sensing matrix can be written as follows

$$A = \frac{1}{\sqrt{m}} \left(\frac{1}{\sqrt{\pi_{K_\ell}}} a_{K_\ell}^* \right)_{1 \leq \ell \leq m}, \tag{9}$$

where $(K_\ell)_{1 \leq \ell \leq m}$ are i.i.d. copies of K such that $\mathbb{P}(K = k) = \pi_k$, for $1 \leq k \leq n$.

We apply Theorem 3.3 when only the degree of sparsity s of the signal to reconstruct is known. This is the setting considered in most CS papers (see e.g. [CT06, Rau10, CP11]). In this context, our main result can be rewritten as follows.

Corollary 4.1. *Let $S \subset \{1, \ldots, n\}$ be a set of indices of cardinality s and suppose that $x \in \mathbb{C}^n$ is an s-sparse vector. Fix $\varepsilon \in (0, 1)$. Suppose that the sampling matrix A is constructed as in* (9)*. If*

$$m \geq C \cdot s \cdot \max_{1 \leq k \leq n} \frac{\|a_k\|_\infty^2}{\pi_k} \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \tag{10}$$

then x is the unique solution of (1) *with probability at least $1 - \varepsilon$.*

Moreover, the drawing distribution minimizing (10) *is $\pi_k = \frac{\|a_k\|_\infty^2}{\sum_{\ell=1}^n \|a_\ell\|_\infty^2}$, which leads to*

$$m \geq C \cdot s \cdot \sum_{k=1}^n \|a_k\|_\infty^2 \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

The proof is given in Appendix D.1.

Note that Corollary 4.1 is identical to Theorem 1.1 in [CP11] up to a logarithmic factor. This result is usually used to explain the practical success of variable density sampling. It is the core of papers such as [PVW11, KW14, CCKW14].

4.2 Isolated measurements with structured sparsity

When using coherent transforms, meaning that the term $\max_{1 \leq k \leq n} \frac{\|a_k\|_\infty^2}{\pi_k}$ in Equation (10) is an increasing function of n, Corollary 4.1 is unsufficient to justify the use of CS in applications. In this section, we show that the proposed results allow justifying the use of CS even in the extreme case where the sensing is performed with the canonical basis.

4.2.1 A toy example: sampling isolated measurements from the Identity matrix and knowing the support S

Suppose that the signal x to reconstruct is S-sparse where $S \subseteq \{1, \ldots, n\}$ is a fixed subset. Consider the highly coherent case where $A_0 = \mathrm{Id}$. All current CS theories would give the same unsatisfactory conclusion: it is not possible to use CS since A_0 is a perfectly coherent transform. Indeed, the bound on the required number of isolated measurements given by standard CS theories [CP11] reads as follows

$$m \geq C \cdot s \cdot \max_{1 \leq k \leq n} \frac{\|e_k^*\|_\infty^2}{\pi_k} \ln(n/\varepsilon) = C \cdot s \cdot \max_{1 \leq k \leq n} \frac{1}{\pi_k} \ln(n/\varepsilon).$$

Without any assumption on the support S, one can choose to draw the measurements uniformly at random, i.e. $\pi_k = 1/n$ for $1 \leq k \leq n$. This particular choice leads to a required number of measurements of the order

$$m \geq C \cdot s \cdot n \ln(n/\varepsilon),$$

which corresponds to fully sampling the acquisition space several times.

Let us now see what conclusion can be drawn with Theorem 3.3.

Corollary 4.2. *Let $S \subseteq \{1, \ldots, n\}$ of cardinality s. Suppose that $x \in \mathbb{C}^n$ is an S-sparse vector. Fix $\varepsilon \in (0,1)$. Suppose that the sampling matrix A is constructed as in* (9) *with $A_0 = \mathrm{Id}$. Set $\pi_k = \frac{\delta_{k,S}}{s}$ for $1 \leq k \leq n$ where $\delta_{k,S} = 1$ if $k \in S$, 0 otherwise. Suppose that*

$$m \quad \geq \quad C \cdot s \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

then x is the unique solution of (1) *with probability at least $1 - \varepsilon$.*

With this new result, $O(s\ln(s)\ln(n))$ measurements are sufficient to reconstruct the signal via a totally coherent. The least amount of measurements necessary to recover x is of order $O(s\ln(s))$, by an argument of coupon collector effect [Fel08, p.262]. Therefore, Corollary 4.2 is near-optimal up to logarithmic factors.

Proof. The result ensues from a direct evaluation of Γ. Indeed,

$$\|e_k e_{k,S}^*\|_{\infty\to\infty} = \max_{1\leq i\leq n} \sup_{\|v\|_\infty \leq 1} |\left\langle e_i, e_k e_{k,S}^* v\right\rangle| = \sup_{\|v\|_\infty\leq 1} |e_{k,S}^* v| = \delta_{k,S},$$

where $\delta_{k,S} = 1$ if $k \in S$, 0 otherwise. Therefore

$$\Theta = \max_{1\leq k\leq n} \frac{\delta_{k,S}}{\pi_k}.$$

Then, we can write that

$$\begin{aligned}\Upsilon(S,\pi) &= \max_{1\leq i\leq n} \sup_{\|v\|_\infty\leq 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* e_k e_{k,S}^* v|^2 = \max_{1\leq i\leq n} \sup_{\|v\|_\infty \leq 1} \frac{|e_{i,S}^* v|^2}{\pi_i} \\ &= \max_{1\leq i\leq n} \frac{\delta_{i,S}}{\pi_i}.\end{aligned}$$

To conclude the proof it suffices to apply Theorem 3.3. ■

4.2.2 Isolated measurements when the degree of sparsity is structured by levels

In this part, we consider a partition of $\{1, \ldots, n\}$ into levels $(\Omega_i)_{i=1,\ldots,N} \subset \{1, \ldots, n\}$ such that

$$\bigsqcup_{1\leq i\leq N} \Omega_i = \{1, \ldots, n\} \qquad \text{and} \qquad |\Omega_i| = N_i.$$

We consider that x is S-sparse with $|S \cap \Omega_i| = s_i$ for $1 \leq i \leq N$ meaning that restricted to the level Ω_i, the signal $P_{\Omega_i} x$ is s_i-sparse. This setting is studied extensively in the recent papers [AHPR13, RHA14, BH14]. Theorem 3.3 provides the following guarantees.

Corollary 4.3. *Let $S \subset \{1, \ldots, n\}$ be a set of indices of cardinality s, such that $|S \cap \Omega_i| = s_i$ for $1 \leq i \leq N$. Suppose that $x \in \mathbb{C}^n$ is an S-sparse vector. Fix $\varepsilon \in (0,1)$. Suppose that the sampling matrix A is constructed as in* (9)*. Set*

$$m \quad \geq \quad C \left(\max_{1\leq k\leq n} \frac{\sum_{\ell=1}^N s_\ell \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k} \right) \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \tag{11}$$

$$m \quad \geq \quad C \left(\max_{1\leq i\leq n} \sup_{\|v\|_\infty\leq 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2 \right) \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \tag{12}$$

then x is the unique solution of (1) *with probability at least $1 - \varepsilon$.*

The proof of Corollary 4.3 is given in Appendix D.2.1. We show in Appendix D.2.2 that a simple analysis leads to results that are nearly equivalent to those in [AHPR13]. It should be noted that the term $\frac{\|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}$ is related to the notion of local coherence defined in [AHPR13]. There are however a few differences making our approach potentially more interesting in the case of isolated measurements:

- Our paper is based on i.i.d. sampling with an arbitrary drawing distribution. This leaves a lot of freedom for generating sampling patterns and optimizing the probability π in order to minimize the upper-bounds (11) and (12). In contrast, the results in [AHPR13] are based on uniform Bernoulli sampling over *fixed* levels. The dependency on the levels is not explicit and it therefore seems complicated to optimize them.
- We can deal with a fixed support S, which enlarges the possibilities for structured sparsity. It is also possible to consider random supports as explained in Proposition 3.6.

4.2.3 Isolated measurements for the Fourier-Haar transform

The bounds in Corollary 4.3 are rather cryptic. They have to be analyzed separately for each sampling strategy. To conclude the discussion on isolated measurements, we provide a practical example with the 1D Fourier-Haar system.

We set $A_0 = \mathcal{F}\phi^*$, where $\mathcal{F} \in \mathbb{C}^{n\times n}$ is the 1D Fourier transform and $\phi^* \in \mathbb{C}^{n\times n}$ is the 1D inverse wavelet transform. To simplify the notation, we assume that $n = 2^J$ and we decompose the signal at the maximum level $J = \log_2(n) - 1$. In order to state our result, we introduce a dyadic partition $(\Omega_j)_{0\le j\le J}$ of the set $\{1,\dots,n\}$. We set $\Omega_0 = \{1\}$, $\Omega_1 = \{2\}$, $\Omega_3 = \{3,4\}, \dots, \Omega_J = \{n/2+1,\dots,n\}$. We also define the function $j : \{1,\dots,n\} \to \{0,\dots,J\}$ by $j(u) = j$ if $u \in \Omega_j$.

Corollary 4.4. *Let $S \subset \{1,\dots,n\}$ be a set of indices of cardinality s, such that $|S\cap\Omega_j| = s_j$ for $0 \le j \le J$. Suppose that $x \in \mathbb{C}^n$ is an s-sparse vector supported on S. Fix $\varepsilon \in (0,1)$. Suppose that A is constructed from the Fourier-Haar transform A_0. Choose π_k to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$. If*

$$m \;\ge\; C \cdot \max_{0\le j\le J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right), \tag{13}$$

then x is the unique solution of (1) with probability at least $1-\varepsilon$.

In particular, the distribution minimizing (13) is

$$\tilde{\pi}_j = \frac{2^{-j}\sum_{p=0}^{J} 2^{-|j-p|/2} s_p}{\sum_{\ell=1}^{n} 2^{-j(\ell)} \sum_{p=0}^{J} 2^{-|j(\ell)-p|/2} s_p},$$

which leads to

$$m \;\ge\; C \cdot \sum_{j=0}^{J} \left(s_j + \sum_{\substack{p=0\\ p\ne j}}^{J} 2^{-|j-p|/2} s_p \right) \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right). \tag{14}$$

The proof is presented in Section D.3. This corollary is once again similar to the results in [AHR14b]. The number of measurements in each level j should depend on the degree of sparsity s_j but also on the degree of sparsity of the other levels which is more and more attenuated when the level is far away from the j-th one.

Remark 4.5. *The Fourier-Wavelet system is coherent and the initial compressed sensing theories cannot explain the success of sampling strategies with such a transform. To overcome the coherence, two strategies have been devised. The first one is based on variable density sampling*

(see e.g. [PMG+12, CCKW14, KW14]). The second one is based on variable density sampling and an additional structured sparsity assumption (see e.g. [AHPR13] and Corollary 4.4). First, note that the results obtained with the latter approach allow recovering signal with arbitrary supports. Indeed, $\sum_{j=0}^{J} s_j + \sum_{\substack{p=0 \\ p\neq j}}^{J} 2^{-|j-p|/2} s_p \leq 2s$.

Second, it is not clear yet - from a theoretical point of view - that the structure assumption allows obtaining better guarantees. Indeed, it is possible to show that the sole variable density sampling leads to perfect reconstruction from $m \propto s\ln(n)^2$ *measurements, which is on par with bound* (14)*. It will become clear that structured sparsity is essential when using the Fourier-Wavelet systems with structured acquisition. Morever, the numerical experiments led in [AHR14b] let no doubt about the fact that structured sparsity is essential to ensure good reconstruction with a low number of measurements.*

4.3 Structured acquisition and structured sparsity

In this paragraph, we illustrate how Theorem 3.3 explains the practical success of structured acquisition in applications. We will mainly focus on the 2D setting: the vector $x \in \mathbb{C}^n$ to reconstruct can be seen as an image of size $\sqrt{n} \times \sqrt{n}$.

4.3.1 The limits of structured acquisition

In [BBW14, PDG15], the authors provided theoretical CS results when using block-constrained acquisitions. Moreover, the results in [BBW14] are proved to be tight in many practical situations. Unfortunately, the bounds on the number of blocks of measurements necessary for perfect reconstruction are incompatible with a faster acquisition.

To illustrate this fact, let us recall a typical result emanating from [BBW14]. It shows that the recovery of sparse vectors with an arbitrary support is of little interest when sampling lines of tensor product transforms. This setting is widely used in imaging. It corresponds to the MRI sampling strategy proposed in [LDP07].

Proposition 4.6 ([BBW14])**.** *Suppose that* $A_0 = \phi \otimes \phi \in \mathbb{C}^{n\times n}$ *is a 2D separable transform, where* $\phi \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ *is an orthogonal transform. Consider blocks of measurements made of* $\sqrt{n}$ *horizontal lines in the 2D acquisition space, i.e. for* $1 \leq k \leq \sqrt{n}$

$$B_k = \left(\phi_{k,1}\phi, \ldots, \phi_{k,\sqrt{n}}\phi\right).$$

If the number of acquired lines m *is less than* $\min(2s, \sqrt{n})$*, then there exists no decoder* Δ *such that* $\Delta(Ax) = x$ *for all* s*-sparse vector* $x \in \mathbb{C}^n$*.*

In other words, the minimal number m *of distinct blocks required to identify every* s*-sparse vectors is necessarily larger than* $\min(2s, \sqrt{n})$*.*

This theoretical bound is quite surprising: it seems to enter in contradiction with the practical results obtained in Figure 2 or with one of the most standard CS strategy in MRI [LDP07]. Indeed, the equivalent number of isolated measurements required by Proposition 4.6 is of the order $O(s\sqrt{n})$. This theoretical result means that in many applications, a full sampling strategy should be adopted, when the acquisition is structured by horizontal lines. In the next paragraphs, we show how Theorem 3.3 allows bridging the gap between theoretical recovery and practical experiments.

4.3.2 Breaking the limits with adapted structured sparsity

In this paragraph, we illustrate - through a simple example - that additional assumptions on structured sparsity is the key to explain practical results.

Corollary 4.7. *Let $A_0 \in \mathbb{C}^{n\times n}$ be the 2D Fourier transform. Assume that x is a 2D signal with support S concentrated on q horizontal lines of the spatial plane, i.e.*

$$S \subset \{(j-1)\sqrt{n} + \{1,\dots,\sqrt{n}\}, j \in J\} \tag{15}$$

where $J \subset \{1,\dots,\sqrt{n}\}$ and $|J| = q$.

Choose a uniform sampling strategy among the $\sqrt{n}$ horizontal lines, i.e. $\pi_k^\star = 1/\sqrt{n}$ for $1 \le k \le \sqrt{n}$. The number m of sampled horizontal lines sufficient to reconstruct x with probability $1-\varepsilon$ is

$$m \ge C \cdot q \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

The proof is given in Appendix D.4 By Proposition 4.7, we can observe that the required number of sampled lines is of the order of non-zero lines in the 2D signal. In comparison, Proposition 4.6 in [BBW14] (with no structured sparsity) requires

$$m \gtrsim s \cdot \ln(n/\varepsilon),$$

measurements, to get the same guarantees. This means that the required number of horizontal lines to sample is of the order of the non-zero coefficients. By putting aside the logarithmic factors, we see that the gain with our new approach is considerable. Clearly, our strategy is able to take advantage of the sparsity structure of the signal of interest.

4.3.3 Consequences for MRI sampling

We now turn to a real MRI application. We assume that the sensing matrix $A_0 \in \mathbb{C}^{n\times n}$ is the product of the 2D Fourier transform $\mathcal{F}_{2D}$ with the inverse 2D wavelet transform Φ^*. We aim at reconstructing a vector $x \in \mathbb{C}^n$ that can be seen as a 2D wavelet transform with $\sqrt{n} \times \sqrt{n}$ coefficients. Set $J = \log_2(\sqrt{n}) - 1$ and let $(\tau_j)_{0\le j\le J}$ denote a dyadic partition of the set $\{1,\dots,\sqrt{n}\}$, i.e. $\tau_0 = \{1\}$, $\tau_1 = \{2\}$, $\tau_2 = \{3,4\}, \dots, \tau_J = \{\sqrt{n}/2+1,\dots,\sqrt{n}\}$. Define $j : \{1,\dots,\sqrt{n}\} \to \{0,\dots,J\}$ by $j(u) = j$ if $u \in \tau_j$. Finally, define the sets $\Omega_{\ell,\ell'} = \tau_\ell \times \tau_{\ell'}$, for $0 \le \ell, \ell' \le J$. See Figure 3 for an illustration of these sets.

Definition 4.8. *Given $S = supp(x)$, define the following quantity*

$$s_\ell^c := \max_{0\le l'\le J} \max_{k\in\tau_{\ell'}} \left|S \cap \Omega_{\ell,\ell'} \cap C_k\right|, \tag{16}$$

where C_k represents the set corresponding to the k-th vertical line (see Figure 3).

The quantity s_ℓ^c represents the maximal sparsity of x restricted to columns (or vertical lines) of $\cup_{1\le l'\le J}\Omega_{\ell,\ell'}$. We have now settled everything to state our result.

As a first step, we will consider the case of Shannon's wavelets, leading to a block-diagonal sampling matrix A_0.

Corollary 4.9. *Let $S \subset \{1,\dots,n\}$ be a set of indices of cardinality s. Suppose that $x \in \mathbb{C}^n$ is an s-sparse vector supported on S. Fix $\varepsilon \in (0,1)$. Suppose that A_0 is the product of the 2D Fourier transform with the 2D inverse Shannon's wavelets transform. Consider that the blocks of measurements are the $\sqrt{n}$ horizontal lines in the 2D setting. Choose $(\pi_k)_{1\le k\le\sqrt{n}}$ to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$. If the number of horizontal lines to acquire satisfies*

$$m \gtrsim \max_{0\le j\le J} \frac{1}{\tilde{\pi}_j} 2^{-j} s_j^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right),$$

then x is the unique solution of Problem 1. Furthermore, choosing $\tilde{\pi}_j = \frac{s_j^c/2^j}{\sum_{\ell=0}^J s_\ell^c}$, for $0 \le j \le J$, leads to the following upper bound

$$m \gtrsim \sum_{j=0}^{J} s_j^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right).$$

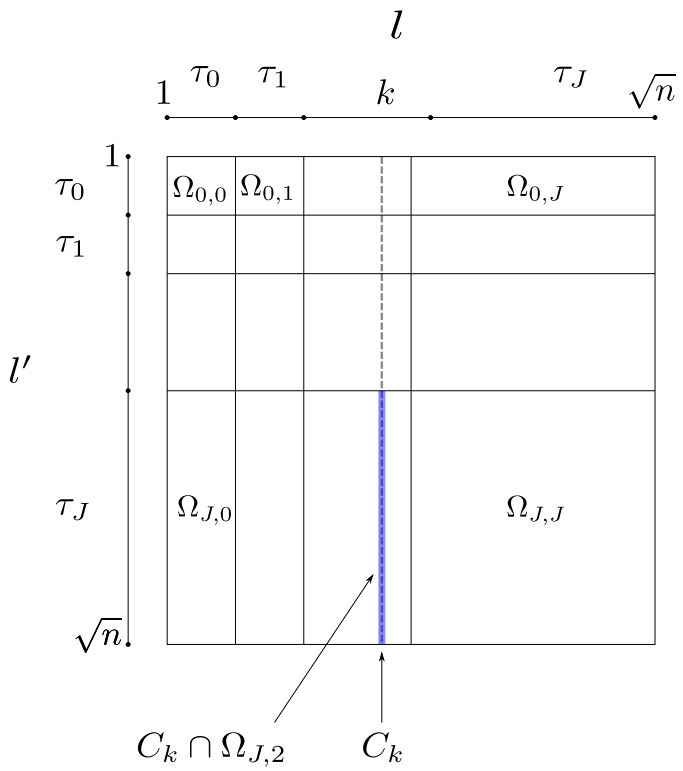


Figure 3: 2D view of the signal $x \in \mathbb{C}^n$ to reconstruct. The vector x can be reshaped in a $\sqrt{n} \times \sqrt{n}$ matrix. C_k represents the coefficient indexes corresponding to the k-th vertical column.

The proof is given in Section D.5. Corollary 4.9 shows that the number of lines acquired at level j depends only on an extra-column structure of S. Now let us turn to a case where the matrix A_0 is not block-diagonal anymore.

Corollary 4.10. *Suppose that $x \in \mathbb{C}^n$ is an S-sparse vector. Fix $\varepsilon \in (0,1)$. Suppose that A_0 is the product of the 2D Fourier transform with the 2D inverse Haar transform. Consider that the blocks of measurements are the $\sqrt{n}$ horizontal lines. Choose $(\pi_k)_{1 \leq k \leq \sqrt{n}}$ to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$.*

If the number m of drawn horizontal lines satisfies

$$m \gtrsim \max_{0 \leq j \leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \ln(s) \ln\left(\frac{n}{\varepsilon}\right),$$

then x is the unique solution of Problem 1 with probablity $1-\varepsilon$.

In particular, if

$$\pi_k = \frac{2^{-j(k)} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c}{\sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c},$$

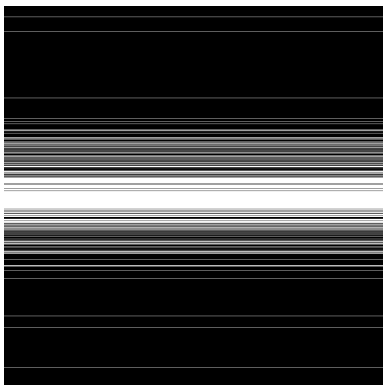
then

$$m \gtrsim \sum_{j=0}^{J} \left(s_j^c + \sum_{\substack{r=0 \\ r \neq j}}^{J} 2^{-|j-r|/2} s_r^c \right) \cdot \ln(s) \ln\left(\frac{n}{\varepsilon}\right)$$

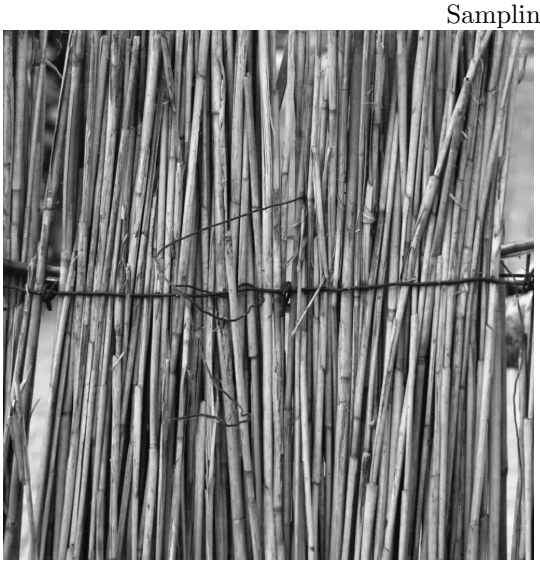
ensures perfect reconstruction with probability $1-\varepsilon$.

The proof of Corollary 4.10 is given in Section D.6.

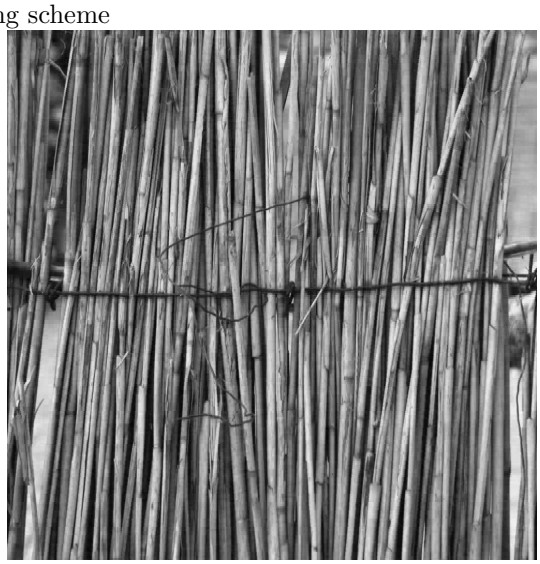
This result indicates that the number of acquired lines in the "horizontal" level j should be chosen depending on the quantities s_j^c. Note that this is very different from the sparsity by levels proposed in [AHPR13]. In conclusion, Corollary 4.10 reveals that with a structured acquisition, the sparsity needs to be more structured in order to guarantee exact recovery. To the best of our knowledge, this is the first theoretical result which can explain why sampling lines in MRI as in [LDP07] might work. In Figure 4, we illustrate that the results in Corollary 4.10 seem to indeed correspond to the practical reality. In this experiment, we seek reconstructing a reeds image from block structured measurements. As a test image, we chose a reeds image with vertical stripes of its rotated version. This particular geometrical structure explains that the quantities s_j^c are much higher for the horizontal stripes than for the vertical one. As can be seen, the image with a low s_j^c is much better reconstructed than the one with a high s_j^c. This wa predicted by our theory.



Sampling scheme

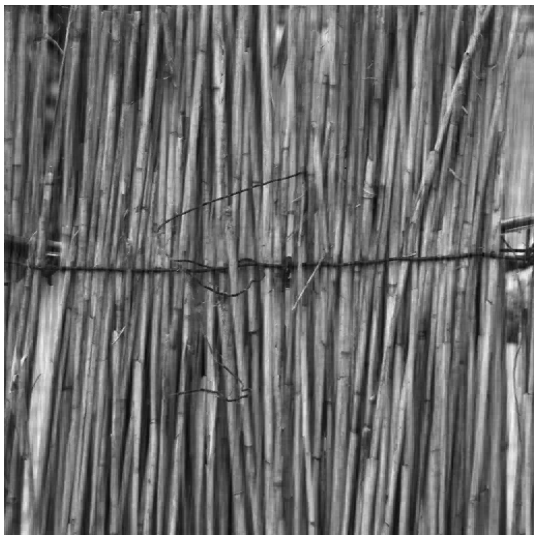


(a) Original image
$s^c = (16, 16, 32, 59, 81, 75, 48)$



(b) SNR = 27.8 dB

(c) Original image
$s^c = (16, 16, 32, 64, 124, 240, 411)$



(d) SNR = 14.7 dB

Figure 4: An example of reconstruction of a 2048×2048 real image sensed in the Fourier domain. In (a) (c), Reference images to reconstruct: (c) is the same image as (a) but rotated of 90°. We precise the value of the vector $s^c = \left(s_j^c\right)_{1 \leq j \leq 7}$ for both images. Note that the quantities s_j^c are larger in the case of image (b). For the reconstruction, we use the sampling scheme at the top of the Figure. It corresponds to 9.8 % of measurements. In (b) (d), corresponding reconstruction via ℓ_1-minimization. We have rotated the image in (d) to facilitate the comparison between both. Note that (b) is much better reconstructed than (d). This is predicted by Corollary 4.10.

5 Extensions

5.1 The case of Bernoulli block sampling

We analyzed the combination of structured acquisition and structured sparsity with i.i.d. drawings of random blocks. These results can be extended to a Bernoulli sampling setting. In such a setting, the sensing matrix is constructed as follows

$$A = \left(\frac{\delta_k}{\sqrt{\pi_k}} B_k \right)_{1 \leq k \leq M},$$

where $(\delta_k)_{1 \leq k \leq M}$ are independent Bernoulli random variables such that $\mathbb{P}(\delta_k = 1) = \pi_k$, for $1 \leq k \leq M$. We may set $\sum_{k=1}^{M} \pi_k = m$ in order to measure m blocks of measurements in expectation. By considering the same definition for $\Gamma(S, \pi)$ with $(\pi_k)_{1 \leq k \leq M}$ the Bernoulli weights, it is possible, for the case of Bernoulli block sampling, to give a reconstruction result that shares a similar flavor to Theorem 3.3.

5.2 Towards new sampling schemes?

The results in Section 4.3.3 lead to the conclusion that exact recovery with structured acquisition can only occur if the the signal to reconstruct possesses an adequate sparsity pattern. We believe that the proposed theorems might help designing new efficient and feasible sampling schemes. Ideally, this could be done by optimizing $\Gamma(S, \pi)$ assuming that S belongs to some set of realistic signals. Unfortunately, this optimization seems unrealistic to perform numerically, owing to the huge dimensions of the objects involved. We therefore leave this question open for future works.

However, probing the limits of a given system, as was proposed in Corollary 4.10 helps designing better sampling schemes. To illustrate this fact, we performed a simple experiment. Since the quantity s_c^j is critical to characterize a sampler efficiency, it is likely that mixing horizontal and vertical sampling lines improves the situation. We aim at reconstructing the MR image shown in Figure 5 and assume that it is sparse in the wavelet basis. In Figure 5(a)(d), we propose two different sampling schemes. The first one is based solely on parallel lines in the horizontal direction, while the second one is based on a combination of vertical and horizontal lines. The combination of vertical and horizontal lines provides much better reconstruction results despite a lower total number of measurements.

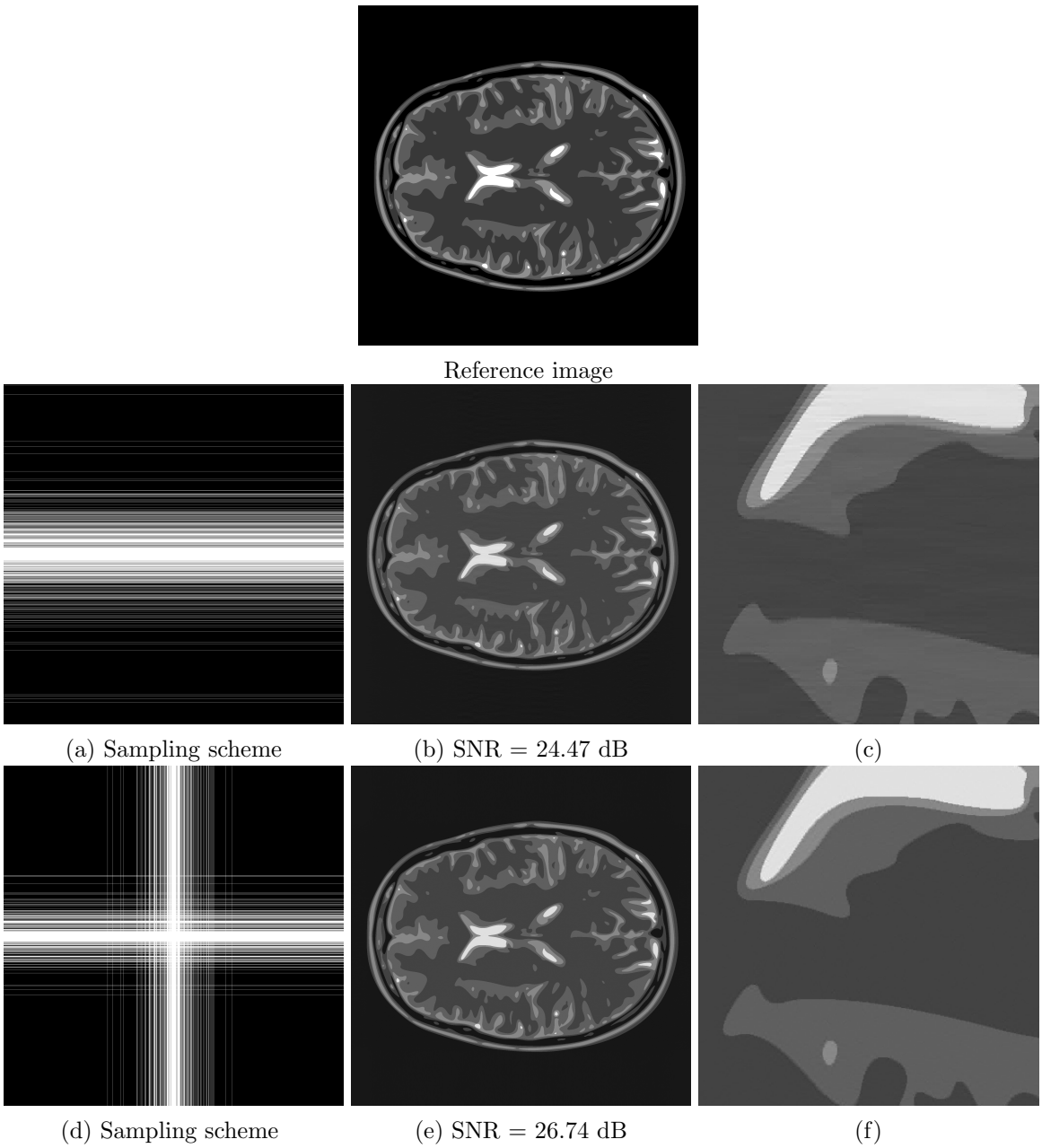


Figure 5: An example of MRI reconstruction of a 2048×2048 phantom. The reference image to reconstruct is presented at the top of the figure. It is considered sparse in the wavelet domain. In (a) (d), we present two kinds of sampling schemes with 20 % of measurements: the samples are acquired in the 2D Fourier domain. In (b) (e), we show the corresponding reconstruction via ℓ_1-minimization. In (c) (f) we enhance the results by zooming on the reconstructed images. Note that the horizontal and vertical sampling scheme produces much better reconstruction results despite a smaller number of measurements since samples are overlapping. Moreover, the acquisition time would be exactly the same for an MRI.

Acknowledgement

The authors would like to thank Ben Adcock and Anders Hansen for their availibility for discussion. They are also grateful to Nicolas Chauffert for discussion. This work was partially supported by the CIMI (Centre International de Mathématiques et d'Informatique) Excellence program.

A Proofs of the main results

A.1 Proof of Theorem 3.3

In this section, we give sufficient conditions to guarantee that the vector x is the unique minimizer of (1), using an inexact dual certificate see [CP11].

Lemma A.1 (Inexact duality [CP11])**.** *Suppose that $x \in \mathbb{R}^n$ is supported on $S \subset \{1, \dots, n\}$. Assume that A_S is full column rank and that*

$$\| (A_S^* A_S)^{-1} \|_{2\to 2} \leq 2 \qquad \textit{and} \qquad \max_{i \in S^c} \|A_S^* A e_i\|_2 \leq 1, \tag{17}$$

where $(A_S^* A_S)^{-1}$ only makes sense on the set span$\{e_i, i \in S\}$. Morever, suppose that there exists $v \in \mathbb{R}^n$ in the row space of A obeying

$$\|v_S - \operatorname{sign}(x_S)\|_2 \leq 1/4 \qquad \textit{and} \qquad \|v_{S^c}\|_\infty \leq 1/4, \tag{18}$$

Then, the vector x is the unique solution of the minimization problem (1)

First, let us focus on Conditions (17). Remark that $A_S^* A_S$ is invertible by assuming that A_S is full column-rank. Moreover,

$$\| (A_S^* A_S)^{-1} \|_{2\to 2} = \left\| \sum_{k=0}^{\infty} (A_S^* A_S - P_S)^k \right\|_{2\to 2} \leq \sum_{k=0}^{\infty} \|A_S^* A_S - P_S\|_{2\to 2}^k .$$

Therefore, if $\|A_S^* A_S - P_S\|_{2\to 2} \leq 1/2$ is satisfied, then $\| (A_S^* A_S)^{-1} \|_{2\to 2} \leq 2$. Moreover, by Lemma C.1, $\| (A_S^* A_S)^{-1} \|_{2\to 2} \leq 2$ with probability at least $1 - \varepsilon$, provided that

$$m \geq \frac{28}{3} \Theta(S, \pi) \ln\left(\frac{2s}{\varepsilon}\right).$$

By definition of $\Gamma(S, \pi)$, the first inequality of Conditions (17) is therefore ensured with probability larger than $1 - \varepsilon$ if

$$m \geq \frac{28}{3} \Gamma(S, \pi) \ln\left(\frac{2s}{\varepsilon}\right). \tag{19}$$

Furthermore, using Lemma C.5, we obtain that

$$\max_{i \in S^c} \|A_S^* A e_i\|_2 \leq 1$$

with probability larger than $1 - \varepsilon$ if

$$m \geq \Theta(S, \pi) \left(1 + 4\sqrt{\ln\left(\frac{n}{\varepsilon}\right)} + 4 \ln\left(\frac{n}{\varepsilon}\right)\right).$$

Again by definition of $\Gamma(S, \pi)$, the second part of Conditions (18) is ensured if $n \geq 3$ and

$$m \geq 9 \Gamma(S, \pi) \ln\left(\frac{n}{\varepsilon}\right). \tag{20}$$

Conditions (18) remain to be verified. The rest of the proof of Theorem 3.3 relies on the construction of a vector v satisfying the conditions described in Lemma A.1 with high probability. To do so, we adapt the so-called golfing scheme introduced by Gross [Gro11] to our setting. More precisely, we will iteratively construct a vector that converges to a vector v satisfying (18) with high probability.

Let us first partition the sensing matrix A into blocks of blocks so that, from now on, we denote by $A^{(1)}$ the first m_1 blocks of A, $A^{(2)}$ the next m_2 blocks, and so on. The L random

matrices $\left\{A^{(\ell)}\right\}_{\ell=1,\ldots,L}$ are independently distributed, and we have that $m = m_1+m_2+\ldots+m_L$. As explained before, $A_S^{(\ell)}$ denotes the matrix $A^{(\ell)}P_S$.

The golfing scheme starts by defining $v^{(0)} = 0$, and then it iteratively defines

$$v^{(\ell)} = \frac{m}{m_\ell} A^{(\ell)*} A_S^{(\ell)} \left(\operatorname{sign}(x) - v^{(\ell-1)}\right) + v^{(\ell-1)}, \tag{21}$$

for $\ell = 1, \ldots, L$, where $\operatorname{sign}(x_i) = 0$ if $x_i = 0$. In the rest of the proof, we set $v = v^{(L)}$. By construction, v is in the row space of A. The main idea of the golfing scheme is then to combine the results from the various Lemmas in Section C with an appropriate choice of L to show that the random vector v satisfies the assumptions of Lemma A.1 with large probability. Using the shorthand notation $v_S^{(\ell)} = P_S v^{(\ell)}$, let us define

$$w^{(\ell)} = \operatorname{sign}(x) - v_S^{(\ell)},\ \ell = 1, \ldots, L,$$

where $x \in \mathbb{C}^n$ is the solution of Problem (1).

From the definition of $v^{(\ell)}$, it follows that, for any $1 \le \ell \le L$,

$$w^{(\ell)} = \left(P_S - \frac{m}{m_\ell} A_S^{(\ell)*} A_S^{(\ell)}\right) w^{(\ell-1)} = \prod_{j=1}^{\ell} \left(P_S - \frac{m}{m_j} A_S^{(j)*} A_S^{(j)}\right) \operatorname{sign}(x), \tag{22}$$

and

$$v = \sum_{\ell=1}^{L} \frac{m}{m_\ell} A^{(\ell)*} A_S^{(\ell)} w^{(\ell-1)}. \tag{23}$$

Note that in particular, $w^{(0)} = \operatorname{sign}(x)$ and $w^{(L)} = \operatorname{sign}(x) - v$. In what follows, it will be shown that the matrices $P_S - \frac{m}{m_\ell} A_S^{(\ell)*} A_S^{(\ell)}$ are contractions and that the norm of the vector $w^{(\ell)}$ decreases geometrically fast with ℓ. Therefore, $v_S^{(\ell)}$ becomes close to $\operatorname{sign}(x_S)$ as ℓ tends to L. In particular, we will prove that $\|w^{(L)}\|_2 \le 1/4$ for a suitable choice of L. In addition, we also show that v satisfies the condition $\|v_{S^c}\|_\infty \le 1/4$. All these conditions will be shown to be satisfied with a large probability (depending on ε).

For all $1 \le \ell \le L$, assume that

$$\left\|w^{(\ell)}\right\|_2 \le r_\ell \left\|w^{(\ell-1)}\right\|_2 \tag{C1-ℓ}$$

$$\left\|\frac{m}{m_\ell}\left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)}\right\|_\infty \le t_\ell \|w^{(\ell-1)}\|_\infty \tag{C2-ℓ}$$

$$\left\|\left(\frac{m}{m_\ell}\left(A_S^{(\ell)}\right)^* A_S^{(\ell)} - P_S\right) w^{(\ell-1)}\right\|_\infty \le t'_\ell \|w^{(\ell-1)}\|_\infty, \tag{C3-ℓ}$$

with

(i) $L = 2 + \left\lceil \frac{\ln(s)}{2\ln 2} \right\rceil$,

(ii) $r_\ell = \frac{1}{2}$, for $\ell = 1, \ldots, L$,

(iii) $t_\ell = t'_\ell = \frac{1}{5}$ for $\ell = 1, \ldots, L$.

Note that using (C1-ℓ), we can write that

$$\|\operatorname{sign}(x_S) - v_S\|_2 = \|w_S^{(L)}\|_2 \le \|\operatorname{sign}(x_S)\|_2 \prod_{\ell=1}^{L} r_\ell \le \sqrt{s} \prod_{\ell=1}^{L} r_\ell \le \frac{\sqrt{s}}{2^L} \le \frac{1}{4}, \tag{24}$$

where the last inequality follows from the previously specified choice on L.

Furthermore, Equation (C2-ℓ) implies that

$$\begin{aligned}
\|v_{S^c}\|_\infty &= \left\| \sum_{\ell=1}^{L} \frac{m}{m_\ell} \left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)} \right\|_\infty \\
&\leq \sum_{\ell=1}^{L} \left\| \frac{m}{m_\ell} \left(A_{S^c}^{(\ell)}\right)^* A_S^{(\ell)} w^{(\ell-1)} \right\|_\infty \\
&\leq \sum_{\ell=1}^{L} t_\ell \left\| w^{(\ell-1)} \right\|_\infty \\
&\leq \sum_{\ell=1}^{L} t_\ell \prod_{j=1}^{\ell-1} t'_j \\
&= \left(\frac{1}{5}\right) \frac{1-(1/5)^L}{1-1/5} \leq \frac{1}{4}.
\end{aligned} \tag{25}$$

Note that in Inequality (25), the control of the operator norms $\infty \to \infty$ avoids the apparition of $\sqrt{s}$ as in the usual golfing scheme of [CP11]. Indeed, in our proof strategy, we have used the fact that $\|w_0\|_\infty = \|\text{sign}(x_S)\|_\infty = 1$, whereas in [CP11] $\|w_0\|_2 = \|\text{sign}(x_S)\|_2 \leq \sqrt{s}$ is involved. This is a key step in the proof, since the absence of the degree of sparsity at this stage allows to derive results depending only on S and not on its cardinality $s = |S|$.

We denote by $p_1(\ell)$, $p_2(\ell)$ and $p_3(\ell)$ the probabilities that the upper bounds (C1-ℓ), (C2-ℓ) and (C3-ℓ) do not hold.

Let us call "failure C" the event in which one of the $3L$ inequalities (C1-ℓ), (C2-ℓ), (C3-ℓ) is not satisfied. Then,

$$\mathbb{P}\left(\text{failure C}\right) \leq \sum_{\ell=1}^{L} \mathbb{P}\left(\text{failure (C1-}\ell\text{)}\right) + \mathbb{P}\left(\text{failure (C2-}\ell\text{)}\right) + \mathbb{P}\left(\text{failure (C3-}\ell\text{)}\right).$$

Therefore a sufficient condition for $\mathbb{P}\left(\text{failure C}\right) \leq \varepsilon$ is $\sum_{\ell=1}^{L} p_1(\ell) + p_2(\ell) + p_3(\ell) \leq \varepsilon$ which holds provided that $p_1(\ell) \leq \varepsilon/3L$, $p_2(\ell) \leq \varepsilon/3L$ and $p_3(\ell) \leq \varepsilon/3L$ for every $\ell = 1, \ldots, L$. By Lemma C.2, condition $p_1(\ell) \leq \varepsilon/3L$ is satisfied if

$$m_\ell \geq 32\Gamma(S,\pi)\left(\ln\left(\frac{3L}{\varepsilon}\right) + \frac{1}{4}\right).$$

By Lemma C.3, condition $p_2(\ell) \leq \varepsilon/3L$ is satisfied if

$$m_\ell \geq 101\Gamma(S,\pi)\ln\left(\frac{12nL}{\varepsilon}\right).$$

By Lemma C.4, condition $p_3(\ell) \leq \varepsilon/3L$ is satisfied if

$$m_\ell \geq 101\Gamma(S,\pi)\ln\left(\frac{12nL}{\varepsilon}\right).$$

Overall, condition

$$m_\ell \geq 101\Gamma(S,\pi)\ln\left(\frac{12nL}{\varepsilon}\right) \tag{26}$$

ensures that (24) and (25) are satisfied with probability $1-\varepsilon$. Condition

$$m = \sum_{\ell=1}^{L} m_\ell \geq 101\left(\frac{\ln(s)}{2\ln(2)} + 3\right)\Gamma(S,\pi)\ln\left(12nL\varepsilon^{-1}\right)$$

will imply (26). The latter condition can be simplified into

$$m \geq 73 \cdot \Gamma(S, \pi) \ln(64s) \left(\ln \left(\frac{9n}{\varepsilon} \right) + \ln \ln(64s) \right). \tag{27}$$

The latter condition ensures that the random vector v, defined by (23), satisfies Assumptions 18 of Lemma A.1 with probability larger than $1 - \varepsilon$.

Hence, we have thus shown that if conditions (19), (20) and (27) are satisfied, then the Assumptions 17 and 18 of Lemma A.1 simultaneously hold with probability larger than $1 - 3\varepsilon$. Note that bound (27) implies (19) and (20).

B Bernstein's inequalities

Theorem B.1 (Scalar Bernstein Inequality)**.** *Let $x_1, \ldots, x_m$ be independent real-valued, zero-mean, random variables such that $|x_\ell| \leq K$ almost surely for every $\ell \in \{1, \ldots, m\}$. Assume that $\mathbb{E}|x_\ell|^2 \leq \sigma_\ell^2$ for $\ell \in \{1, \ldots, m\}$. Then for all $t > 0$,*

$$\mathbb{P}\left(\left| \sum_{\ell=1}^{m} x_\ell \right| \geq t \right) \leq 2 \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3} \right),$$

with $\sigma^2 \geq \sum_{\ell=1}^{m} \sigma_\ell^2$.

Theorem B.2 (Vector Bernstein Inequality (V1))**.** *[CP11, Theorem 2.6] Let $(y_k)_{1 \leq k \leq m}$ be a finite sequence of independent random complex vectors of dimension n. Suppose that $\mathbb{E} y_k = 0$ and $\|y_k\|_2 \leq K$ a.s. for some constant $K > 0$ and set $\sigma^2 \geq \sum_k \mathbb{E}\|y_k\|_2^2$. Let $Z = \|\sum_{k=1}^{m} y_k\|_2$. Then, for any $0 < t \leq \sigma^2/K$, we have that*

$$\mathbb{P}(Z \geq t) \leq \exp\left(-\frac{(t/\sigma - 1)^2}{4} \right) \leq \exp\left(-\frac{t^2}{8\sigma^2} + \frac{1}{4} \right).$$

Theorem B.3 (Bernstein Inequality for self-adjoint matrices)**.** *Let $(Z_k)_{1 \leq k \leq n}$ be a finite sequence of independent, random, self-adjoint matrices of dimension d, and let a_k be a sequence of fixed self-adjoint matrices. Suppose that Z_k is such that $\mathbb{E} Z_k = 0$ and $\|Z_k\|_{2 \to 2} \leq K$ a.s. for some constant $K > 0$ that is independent of k. Moreover, assume that $\mathbb{E} Z_k^2 \preceq A_k^2$ for each $1 \leq k \leq n$. Define*

$$\sigma^2 = \left\| \sum_{k=1}^{n} A_k^2 \right\|_{2 \to 2}$$

Then, for any $t > 0$, we have that

$$\mathbb{P}\left(\left\| \sum_{k=1}^{n} Z_k \right\|_{2 \to 2} \geq t \right) \leq d \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3} \right).$$

Proof. This result is as an application of the techniques developed in [Tro12] to obtain tail bounds for sum of random matrices. Our arguments follow those in the proof of Theorem 6.1 in [Tro12]. We assume that $K = 1$ since the general result follows by a scaling argument. Using the assumption that $\mathbb{E} Z_k^2 \preceq A_k^2$, and by applying the arguments in the proof of Lemma 6.7 in [Tro12], we obtain that

$$\mathbb{E} \exp(\theta Z_k) \preceq \exp\left(g(\theta) A_k^2 \right),$$

for any real $\theta > 0$, where $g(\theta) = e^\theta - \theta - 1$, and the notation $\exp(A)$ denotes the matrix exponential of a self-adjoint matrice A (see [Tro12] for further details). Therefore, by Corollary 3.7 in [Tro12], it follows that

$$\mathbb{P}\left(\left\| \sum_{k=1}^{n} Z_k \right\|_{2 \to 2} \geq t \right) \leq d \inf_{\theta > 0} \left\{ e^{-\theta t + \sigma^2 g(\theta)} \right\}, \tag{28}$$

where $\sigma^2 = \left\|\sum_{k=1}^n A_k^2\right\|_{2\to 2}$. To conclude, we follow the proof of Theorem 6.1 in [Tro12]. The function $\theta \mapsto -\theta t + \sigma^2 g(\theta)$ attains its minimum for $\theta = \ln(1 + t/\sigma^2)$, which implies that the minimal value of the right-hand size of Inequality (28) is $d \exp\left(-\sigma^2 h(t/\sigma^2)\right)$ where $h(u) = (1+u)\ln(1+u) - u$ for $u \geq 0$. To complete the proof, it suffices to use the standard lower bound $h(u) \geq \frac{u^2/2}{1+u/3}$ for $u \geq 0$. ■

C Estimates: auxiliary results

Let S be the support of the signal to be reconstructed such that $|S| = s$. We set

$$\Lambda(S,\pi) := \max_{1\leq k\leq M} \frac{1}{\pi_k} \left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2}.$$

Note that $\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2} \leq \left\|B_{k,S}^* B_{k,S}\right\|_{\infty\to\infty} \leq \left\|B_k^* B_{k,S}\right\|_{\infty\to\infty}$, therefore,

$$\Lambda(S,\pi) \leq \Theta(S,\pi).$$

To make the notation less cluttered, we will write Λ, Θ, Υ and Γ instead of $\Lambda(S,\pi)$, $\Theta(S,\pi)$, $\Upsilon(S,\pi)$ and $\Gamma(S,\pi)$.

Lemma C.1. *Let $S \subset \{1,\dots,n\}$ be of cardinality of s. Suppose that $\Theta \geq 1$. Then, for any $\delta > 0$, one has that*

$$\mathbb{P}\left(\|A_S^* A_S - P_S\|_{2\to 2} \geq \delta\right) \leq 2s \exp\left(-\frac{m\delta^2/2}{\Theta(1+\delta/3)}\right). \tag{E1}$$

Proof. We decompose the matrix $A_S^* A_S - P_S$ as

$$A_S^* A_S - P_S = \frac{1}{m}\sum_{k=1}^m \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S = \frac{1}{m}\sum_{k=1}^m X_k,$$

where $X_k := \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right)$. It is clear that $\mathbb{E}X_k = 0$, and since for all $1 \leq k \leq M$, $\frac{\left\|B_{k,S} B_{k,S}^*\right\|_{2\to 2}}{\pi_k} \leq \Lambda \leq \Theta$, we have that

$$\|X_k\|_{2\to 2} \leq \max\left(\frac{\left\|B_{J_k,S}^* B_{J_k,S}\right\|_{2\to 2}}{\pi_{J_k}} - 1, 1\right) \leq \Theta.$$

Lastly, we remark that

$$\begin{aligned} 0 \preceq \mathbb{E}X_k^2 = \mathbb{E}\left[\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}}\right]^2 - P_S &\preceq \max_{1\leq k\leq M} \frac{\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2}}{\pi_k} \mathbb{E}\left[\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}}\right] \\ &\preceq \max_{1\leq k\leq M} \frac{\left\|B_{k,S}^* B_{k,S}\right\|_{2\to 2}}{\pi_k} P_S \preceq \Lambda P_S \\ &\preceq \Theta P_S. \end{aligned}$$

Therefore, using Theorem B.3, we can set $\sigma^2 = \left\|\sum_{k=1}^m \mathbb{E}X_k^2\right\|_{2\to 2} \leq m\Theta$. Hence, inequality (E1) immediately follows from Bernstein's inequality for random matrices (see Theorem B.3). ■

Lemma C.2. *Let $S \subset \{1,\dots,n\}$, such that $|S| = s$. Let w be a vector in $\mathbb{C}^n$. Then, for any $0 \leq t \leq 1$, one has that*

$$\mathbb{P}\left(\|(A_S^* A_S - P_S)\, w\|_2 \geq t\|w\|_2\right) \leq \exp\left(-\frac{mt^2}{8\Theta} + \frac{1}{4}\right). \tag{E2}$$

Proof. Without loss of generality we may assume that $\|\boldsymbol{w}\|_2 = 1$. We remark that

$$(A_S^* A_S - \mathrm{Id}_s)\, w_S = \frac{1}{m}\sum_{k=1}^{m}\left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) w = \frac{1}{m}\sum_{k=1}^{m} y_k,$$

where $y_k = \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) w$ is a random vector with zero mean. Simple calculations yield that

$$\begin{aligned}
\left\|\frac{1}{m} y_k\right\|_2^2 &= \frac{1}{m^2}\left(w^*\left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}}\right)^2 w - 2w^*\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w + w^* w\right) \\
&\le \frac{1}{m^2}\left(\Lambda w^*\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w - 2w^*\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w + 1\right) \\
&= \frac{1}{m^2}\left((\Lambda - 2)\, w^*\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} w + 1\right) \\
&\le \frac{1}{m^2}\left((\Lambda - 2)\,\Lambda\|w\|_2^2 + 1\right) = \frac{1}{m^2}\left((\Lambda - 2)\,\Lambda + 1\right) \\
&\le \frac{1}{m^2}(\Lambda - 1)^2 \le \frac{1}{m^2}\Lambda^2 \le \frac{1}{m^2}\Theta^2.
\end{aligned}$$

Now, let us define $Z = \left\|\frac{1}{m}\sum_{k=1}^{m} y_k\right\|_2$. By independence of the random vectors y_k, it follows that

$$\begin{aligned}
\mathbb{E}\left[Z^2\right] &= \frac{1}{m}\mathbb{E}\,\|y_1\|_2^2 = \frac{1}{m}\mathbb{E}\left[\left\langle \frac{B_{J,S}^* B_{J,S}}{\pi_J} w, \frac{B_{J,S}^* B_{J,S}}{\pi_J} w\right\rangle - 2\left\langle \frac{B_{J,S}^* B_{J,S}}{\pi_J} w, w\right\rangle + \langle w, w\rangle\right] \\
&= \frac{1}{m}\mathbb{E}\left[\left\langle \left(\frac{B_{J,S}^* B_{J,S}}{\pi_J}\right)^2 w, w\right\rangle - 2\frac{\|B_{J,S} w\|_2^2}{\pi_J} + 1\right].
\end{aligned}$$

To bound the first term in the above equality, one can write

$$\begin{aligned}
\mathbb{E}\left[\left\langle \left(\frac{B_{J_1,S}^* B_{J_1,S}}{\pi_{J_1}}\right)^2 w, w\right\rangle\right] &= \left\langle \mathbb{E}\left[\left(\frac{B_{J_1,S}^* B_{J_1,S}}{\pi_{J_1}}\right)^2\right] w, w\right\rangle \\
&\le \Lambda\left\langle \mathbb{E}\left[\left(\frac{B_{J_1,S}^* B_{J_1,S}}{\pi_{J_1}}\right)\right] w, w\right\rangle \le \Lambda\|w\|_2^2 \le \Theta.
\end{aligned}$$

One immediately has that $\mathbb{E}\frac{\|B_{J,S} w\|_2^2}{\pi_k} = \|w\|_2^2 = 1$. Therefore, one finally obtains that

$$\mathbb{E}\left[Z^2\right] \le \frac{\Theta - 1}{m} \le \frac{\Theta}{m}.$$

Using the above upper bounds, namely $\left\|\frac{1}{m} y_k\right\|_2 \le \frac{\Theta}{m}$ and $\mathbb{E}\left[Z^2\right] \le \frac{\Theta}{m}$, the result of the lemma is thus a consequence of the Bernstein's inequality for random vectors (see Theorem B.2), which completes the proof. ■

Lemma C.3. *Let $S \subset \{1, \ldots, n\}$, such that $|S| = s$. Let v be a vector of $\mathbb{C}^n$. Then we have*

$$\mathbb{P}\left(\|A_{S^c}^* A_S v\|_\infty \ge t\|v\|_\infty\right) \le 4n \exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3}\right). \tag{E3}$$

Proof. Suppose without loss of generality that $\|v\|_\infty = 1$. Then,

$$\|A_{S^c}^* A_S v\|_\infty = \max_{i \in S^c} |\langle e_i, A^* A_S v\rangle| = \max_{i \in S^c} \frac{1}{m}\left|\sum_{k=1}^{m}\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right|.$$

Let us define $Z_k = \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle$. Note that $\mathbb{E}Z_k = 0$, since for $i \in S^c$, $\mathbb{E}\left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle = e_i^* \sum_{k=1}^M \pi_k \frac{B_k^* B_{k,S}}{\pi_k} v = e_i^* P_S v = 0$. From Holder's inequality, we get

$$
\begin{aligned}
|Z_k| &= \left| \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle \right| = \left| e_i^* \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right| \leq \max_{\substack{j \in S^c \\ 1 \leq k \leq M}} \frac{1}{\pi_k} \left\| B_{k,S}^* B_k e_j \right\|_1 \|v\|_\infty \\
&\leq \max_{1 \leq k \leq n} \max_{\substack{j \in S^c \\ 1 \leq k \leq M}} \frac{1}{\pi_k} \left\| e_j^* B_k^* B_{k,S} \right\|_1 = \Theta.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\mathbb{E}|Z_k|^2 &= \mathbb{E} \left| \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle \right|^2 = \sum_{\ell=1}^M \frac{|e_i^* B_\ell^* B_{\ell,S} v|^2}{\pi_\ell} \\
&\leq \Upsilon.
\end{aligned}
$$

Therefore $\sum_{k=1}^m \mathbb{E}|Z_k|^2 \leq m\Upsilon$. Using real-valued Bernstein's inequality B.1 in the case of complex random variables, we obtain

$$
\begin{aligned}
&\mathbb{P}\left(\frac{1}{m} \left| \sum_{k=1}^m \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle \right| \geq t \right) \\
&\leq \mathbb{P}\left(\frac{1}{m} \left| \sum_{k=1}^m \mathrm{Re} \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle \right| \geq t/\sqrt{2} \right) ... \\
&\quad + \mathbb{P}\left(\frac{1}{m} \left| \sum_{k=1}^m \mathrm{Im} \left\langle e_i, \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}} v \right\rangle \right| \geq t/\sqrt{2} \right) \\
&\leq 4 \exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3} \right).
\end{aligned}
$$

Taking the union bound over $i \in S^c$ completes the proof. ■

Lemma C.4. *Let $S \subset \{1, \ldots, n\}$, such that $|S| = s$. Suppose that $\Theta \geq 1$.Let v be a vector of $\mathbb{C}^n$. Then we have*

$$
\mathbb{P}\left(\|(A_S^* A_S - P_S) v\|_\infty \geq t\|v\|_\infty \right) \leq 4s \exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3} \right). \tag{E4}
$$

Proof. Suppose without loss of generality that $\|v\|_\infty = 1$. Then,

$$
\|(A_S^* A_S - P_S) v\|_\infty = \max_{i \in S} |\langle e_i, (A_S^* A_S - P_S) v \rangle| = \max_{i \in S} \frac{1}{m} \left| \sum_{k=1}^m \left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S \right) v \right\rangle \right|.
$$

Let us define $Z_k = \left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S \right) v \right\rangle$. Note that $\mathbb{E}Z_k = 0$. From Holder's inequality, we get

$$
|Z_k| = \left| \left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S \right) v \right\rangle \right| \leq \left\| \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S \right\|_{\infty \to \infty} \leq \max(\Theta - 1, 1) \leq \Theta,
$$

since $\|B_{k,S}^* B_{k,S}\|_{\infty\to\infty} \le \|B_k^* B_{k,S}\|_{\infty\to\infty}$, and using the same argument as in Lemma C.3. Furthermore,

$$\begin{aligned}
\mathbb{E}|Z_k|^2 &= \mathbb{E}\left|\left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) v\right\rangle\right|^2 \\
&= \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right|^2 - \langle e_i, v\rangle\, \mathbb{E}\left\langle e_i, \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle - \langle e_i, v\rangle^*\, \mathbb{E}\left\langle e_i, \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle \\
&\quad + |\langle e_i, v\rangle|^2 \\
&= \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right|^2 - |\langle e_i, v\rangle|^2 \le \mathbb{E}\left|\left\langle e_i, \frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} v\right\rangle\right|^2 = \sum_{\ell=1}^{M} \frac{\left|e_i^* B_{\ell,S}^* B_{\ell,S} v\right|^2}{\pi_\ell} \\
&\le \Upsilon.
\end{aligned}$$

Therefore, $\sum_{k=1}^m \mathbb{E}|Z_k|^2 \le m\Upsilon$, and using real-valued Bernstein's inequality B.1 in the case of complex random variables, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) v\right\rangle\right| \ge t\right) \\
&\le \mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \mathrm{Re}\left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) v\right\rangle\right| \ge t/\sqrt{2}\right) + \mathbb{P}\left(\frac{1}{m}\left|\sum_{k=1}^m \mathrm{Im}\left\langle e_i, \left(\frac{B_{J_k,S}^* B_{J_k,S}}{\pi_{J_k}} - P_S\right) v\right\rangle\right| \ge t/\sqrt{2}\right) \\
&\le 4\exp\left(-\frac{mt^2/4}{\Upsilon + \Theta t/3}\right).
\end{aligned}$$

Taking the union bound over $i \in S$ completes the proof. ■

Lemma C.5. *Let S be a subset of $\{1, \dots, n\}$. Then, for any $0 \le t \le m$, one has that*

$$\mathbb{P}\left(\max_{i\in S^c} \|A_S^* A e_i\|_2 \ge t\right) \le n \exp\left(-\frac{\left(\sqrt{m}t/\sqrt{\Theta} - 1\right)^2}{4}\right). \tag{E5}$$

Proof. Let us fix some $i \in S^c$. For $k = 1, \dots, M$, we define the random vector

$$x_k := \frac{B_{J_k,S}^* B_{J_k}}{\pi_{J_k}} e_i.$$

Then, since $i \in S^c$ one easily gets $\mathbb{E}x_k = \sum_{\ell=1}^M B_{\ell,S}^* B_\ell e_i = \sum_{\ell=1}^M (B_\ell P_S)^* B_\ell e_i = P_S \sum_{\ell=1}^M B_\ell^* B_\ell e_i = P_S e_i = 0$ (note that P_S is self-adjoint). In addition, we can write

$$\|A_S^* A e_i\|_2 = \left\|\frac{1}{m}\sum_{k=1}^m \frac{B_{J_k,S}^* B_{J_k}}{\pi_{J_k}} e_i\right\|_2 = \left\|\frac{1}{m}\sum_{k=1}^M x_k\right\|_2.$$

Then,

$$\begin{aligned}
\|x_k\|_2 &= \left\|\frac{B_{J_k,S}^* B_{J_k}}{\pi_{J_k}} e_i\right\|_2 \le \left\|\frac{B_{J_k,S}^* B_{J_k}}{\pi_{J_k}} e_i\right\|_1 = \left\|e_i^* \frac{B_{J_k}^* B_{J_k,S}}{\pi_{J_k}}\right\|_1 \\
&\le \frac{1}{\pi_{J_k}} \left\|B_{J_k}^* B_{J_k,S}\right\|_{\infty\to\infty} \le \Theta.
\end{aligned}$$

Furthermore, one has that

$$\mathbb{E}\,\|x_k\|_2^2 = \mathbb{E}\left\|\frac{B_{J_k,S}^* B_{J_k}}{\pi_{J_k}} e_i\right\|_2^2 \le \mathbb{E}\left\|\frac{B_{J_k,S}}{\sqrt{\pi_{J_k}}}\right\|_{2\to 2}^2 \left\|\frac{B_{J_k}}{\sqrt{\pi_{J_k}}} e_i\right\|_2^2 \le \Lambda \mathbb{E}\left\|\frac{B_{J_k}}{\sqrt{\pi_{J_k}}} e_i\right\|_2^2 = \Lambda \|e_i\|_2^2 = \Lambda,$$

$$\sum_{k=1}^m \mathbb{E}\,\|x_k\|_2^2 \le m\Lambda \le m\Theta.$$

Hence, using the above upper bounds, it follows from Bernstein's inequality for random vectors (see Theorem B.2) that

$$\mathbb{P}\left(\|A_S^* A e_i\|_2 \geq t\right) \leq \exp\left(-\frac{\left(\sqrt{m}t/\sqrt{\Theta}-1\right)^2}{4}\right),$$

Finally, Inequality (E4) follows from a union bound over $i \in S^c$, which completes the proof. ■

D Proof of results in Applications

D.1 Proof of Corollary 4.1

The proof relies on the evaluation of Θ and Υ in the case of isolated measurements. In this case, we have n blocks composed of isolated measurements. Then, each block corresponds to one of the rows $(a_k^*)_{1\leq k\leq n}$ of A_0. Recall that $\|a_k a_{k,S}^*\|_{\infty\to\infty} = \max_{1\leq i\leq n}\sup_{\|v\|_\infty\leq 1}|e_i^* a_k a_{k,S}^* v|$, so the norm $\|a_k a_{k,S}^*\|_{\infty\to\infty}$ is the maximum ℓ_1-norm of the rows of the matrix $a_k a_{k,S}^*$. Therefore, the quantities in Definition 3.1 can be rewritten as follows

$$\begin{aligned}
\Theta(S,\pi) &:= \max_{1\leq k\leq n}\frac{\|a_k a_{k,S}^*\|_{\infty\to\infty}}{\pi_k} \\
&= \max_{1\leq k\leq n}\frac{\|a_k\|_\infty\|a_{k,S}\|_1}{\pi_k} \qquad (29)\\
&\leq s\cdot\max_{1\leq k\leq n}\frac{\|a_k\|_\infty^2}{\pi_k},
\end{aligned}$$

$$\begin{aligned}
\Upsilon(S,\pi) &= \max_{1\leq i\leq n}\sup_{\|v\|_\infty\leq 1}\sum_{k=1}^n\frac{1}{\pi_k}|e_i^* a_k|^2|a_{k,S}^* v|^2 \qquad (30)\\
&\leq \sup_{\|v\|_\infty\leq 1}\sum_{k=1}^n\frac{1}{\pi_k}\|a_k\|_\infty^2|a_{k,S}^* v|^2 \\
&\leq \sup_{\|v\|_\infty\leq 1}\max_{1\leq \ell\leq n}\frac{\|a_\ell\|_\infty^2}{\pi_\ell}\sum_{k=1}^n|a_{k,S}^* v|^2 = \sup_{\|v\|_\infty\leq 1}\|A_0 P_S v\|_2^2\max_{1\leq\ell\leq n}\frac{\|a_\ell\|_\infty^2}{\pi_\ell} \\
&= \sup_{\|v\|_\infty\leq 1}\|P_S v\|_2^2\max_{1\leq\ell\leq n}\frac{\|a_\ell\|_\infty^2}{\pi_\ell} \\
&\leq s\cdot\max_{1\leq k\leq n}\frac{\|a_k\|_\infty^2}{\pi_k}.
\end{aligned}$$

Therefore we can choose $\Gamma(S,\pi) = s\cdot\max_{1\leq k\leq n}\frac{\|a_k\|_\infty^2}{\pi_k}$, and the result follows by Theorem 3.3.

D.2 Around Corollary 4.3

D.2.1 Proof of Corollary 4.3

Again, this is all about evaluating Θ and Υ in this specific case. Concerning the evaluation of Υ, we can use the expression (30) to conclude that

$$\Upsilon(S,\pi) = \max_{1\leq i\leq n}\sup_{\|v\|_\infty\leq 1}\sum_{k=1}^n\frac{1}{\pi_k}|e_i^* a_k|^2|a_{k,S}^* v|^2.$$

To control Θ, using (29), it suffices to write:

$$\Theta(S,\pi) = \max_{1\le k\le n} \frac{\|a_k a_{k,S}^*\|_{\infty\to\infty}}{\pi_k} \le \max_{1\le k\le n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k}$$
$$\le \max_{1\le k\le n} \frac{\|a_k\|_\infty \sum_{\ell=1}^N \|a_{k,\Omega_\ell}\|_\infty s_\ell}{\pi_k}.$$

By Theorem 3.3, the two conditions

$$m \ \ge\ C\left(\max_{1\le k\le n} \frac{\sum_{\ell=1}^N s_\ell \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}\right) \ln(s)\ln\left(\frac{n}{\varepsilon}\right),$$
$$m \ \ge\ C\left(\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2\right) \ln(s)\ln\left(\frac{n}{\varepsilon}\right),$$

lead to the desired conclusion.

D.2.2 Comparison of Corollary 4.3 and the results in [AHPR13].

Note that the sampling in [AHPR13] is based on Bernoulli drawings structured by level. Their results are then easily transposable to the case of i.i.d. sampling with constant probability by level. The first condition on m in Corollary 4.3 is similar to condition (4.4) in Theorem 4.4 of [AHPR13], since we recognize the term $\frac{\|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty}{\pi_k}$ as the (k,ℓ)-local coherence defined in [AHPR13]. Let us show that the second condition on m is similar to equation (4.5) in [AHPR13]. First, observe that

$$\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2 \le \max_{1\le \ell\le N} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty^2 \left|a_{k,S}^* v\right|^2$$
$$\le \max_{1\le \ell\le N} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty \left|a_{k,S}^* v\right|^2.$$

Let $\tilde{v}$ denote the maximizer in the last expression, and define $\widetilde{s_k} = \left|a_{k,S}^* \tilde{v}\right|^2$ for $1 \le k \le n$. It follows,

$$\max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^n \frac{1}{\pi_k} |e_i^* a_k|^2 \left|a_{k,S}^* v\right|^2 \le \max_{1\le \ell\le N} \sum_{k=1}^n \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty \|a_k\|_\infty \widetilde{s_k}, \tag{31}$$

and $\sum_{k=1}^n \widetilde{s_k} = \sum_{k=1}^n \left|a_{k,S}^* v\right|^2 = \|A_0 P_S \widetilde{v}\|_2^2 = \|P_S \widetilde{v}\|_2^2 \le \sum_{\ell=1}^N s_\ell$. The last inequality and Equation (31) for i.i.d sampling correspond to the condition (4.5) in Theorem 4.4 of [AHPR13] in the case of Bernoulli sampling. This completes the comparison between Corollary 4.3 and the results in [AHPR13].

D.3 Proof of Corollary 4.4

Recall that $(\Omega_j)_{0\le j\le J}$ the dyadic partition of the set of indexes $\{1,\dots,n\}$. Recall also the function $j : \{1,\dots,n\} \to \{0,\dots,J\}$ defined by $j(u) = j$ if $u \in \Omega_j$. In the interests of simplifying notation, in this section, the symbol '$\gtrsim$' will be equivalent to '$\ge C\cdot$', with C a universal constant. The following lemma will be useful to bound above the coefficients of A_0 in absolute value, and to derive Lemmas D.2 and D.3.

Lemma D.1. *[AHR14a] The magnitude of the coefficients of matrix $A_0 = \mathcal{F}\phi^*$, where $\mathcal{F}$ is the 1D Fourier transform and ϕ is the 1D Haar transform, satisfies*

$$\|P_{\Omega_j} A_0 P_{\Omega_\ell}\|_{1\to\infty}^2 \lesssim 2^{-j} 2^{-|j-\ell|}, \quad \textit{for} \quad 0 \le j,\ell \le J. \tag{32}$$

Lemma D.2. *In the case of isolated measurements, with $A_0 = \mathcal{F}\phi^*$ with ϕ to be the inverse 1D Haar transform, suppose that the signal to reconstruct x is sparse by levels, meaning that $\|P_{\Omega_j} x\|_0 \leq s_j$ for $0 \leq j \leq J$. Then,*

$$\Theta \lesssim \max_{1\leq k\leq n} \frac{2^{-j(k)}}{\pi_k} \left(s_{j(k)} + \sum_{\substack{\ell=0 \\ \ell\neq j(k)}}^{J} s_\ell 2^{-|j(k)-\ell|/2} \right). \tag{33}$$

Choosing π_k to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$, the last expression can be rewritten as follows

$$\Theta \lesssim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \left(s_j + \sum_{\substack{\ell=0 \\ \ell\neq j}}^{J} s_\ell 2^{-|j-\ell|/2} \right). \tag{34}$$

Proof. Using (29), we can write

$$\begin{aligned}
\Theta &= \max_{1\leq k\leq n} \frac{\|a_k\|_\infty \|a_{k,S}\|_1}{\pi_k} \leq \max_{1\leq k\leq n} \frac{\|a_k\|_\infty \sum_{\ell=0}^{J} \|a_{k,\Omega_\ell}\|_\infty s_\ell}{\pi_k} \\
&\lesssim \max_{1\leq k\leq n} \frac{1}{\pi_k} 2^{-j(k)/2} \sum_{\ell=0}^{J} 2^{-j(k)/2} 2^{-|j(k)-\ell|/2} s_\ell \\
&\lesssim \max_{1\leq k\leq n} \frac{1}{\pi_k} 2^{-j(k)} \sum_{\ell=0}^{J} 2^{-|j(k)-\ell|/2} s_\ell,
\end{aligned}$$

where we use (32) to bound above $\|a_{k,\Omega_\ell}\|_\infty$. ■

Lemma D.3. *In the case of isolated measurements, with $A_0 = \mathcal{F}\phi^*$ with ϕ to be the inverse Haar transform, suppose that the signal to reconstruct x is sparse by levels, meaning that $\|P_{\Omega_j} x\|_0 \leq s_j$ for $0 \leq j \leq J$. Choosing π_k to be constant by level, i.e. $\pi_k = \tilde{\pi}_{j(k)}$, we have*

$$\Upsilon \lesssim \max_{0\leq j\leq J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p. \tag{35}$$

Proof. Denoting $\tilde{v} = \tilde{v}(i)$ the argument of the supremum in the definition of Υ, we get

$$\begin{aligned}
\Upsilon &:= \max_{1\leq i\leq n} \sum_{k=1}^{n} \frac{1}{\pi_k} |e_i^* a_k|^2 |a_{k,S}\tilde{v}|^2 \leq \max_{0\leq \ell\leq J} \sum_{k=1}^{n} \frac{1}{\pi_k} \|a_{k,\Omega_\ell}\|_\infty^2 |a_{k,S}\tilde{v}|^2 \\
&\lesssim \max_{0\leq \ell\leq J} \sum_{k=1}^{n} \frac{1}{\pi_k} 2^{-j(k)} 2^{-|j(k)-\ell|} |a_{k,S}\tilde{v}|^2 \lesssim \max_{0\leq \ell\leq J} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} 2^{-j} 2^{-|j-\ell|} \underbrace{\sum_{k\in\Omega_j} |a_{k,S}\tilde{v}|^2}_{=:K_j}
\end{aligned}$$

We can rewrite K_j as follows $K_j = \|P_{\Omega_j} A_0 P_S \tilde{v}\|_2^2$. Therefore, since $\|\tilde{v}\|_\infty \leq 1$,

$$\begin{aligned}
\sqrt{K_j} &= \|P_{\Omega_j} A_0 P_S \tilde{v}\|_2 = \|P_{\Omega_j} A_0 \sum_{p=0}^{J} P_{\Omega_p} P_S \tilde{v}\|_2 \leq \sum_{p=0}^{J} \|P_{\Omega_j} A_0 P_{\Omega_p} P_S \tilde{v}\|_2 \\
&\leq \sum_{p=0}^{J} \|P_{\Omega_j} A_0 P_{\Omega_p}\|_{2\to 2} \|P_{\Omega_p} P_S \tilde{v}\|_2 \leq \sum_{p=0}^{J} \|P_{\Omega_j} A_0 P_{\Omega_p}\|_{2\to 2} \sqrt{s_p}.
\end{aligned}$$

Using Lemma 4.3 of [AHR14a], we have the following upper bound

$$\|P_{\Omega_j} A_0 P_{\Omega_p}\|_{2\to 2} \lesssim 2^{-|j-p|/2}, \qquad \text{for} \quad 0 \le j, p \le J.$$

Then, $\sqrt{K_j} \lesssim \sum_{p=0}^{J} 2^{-|j-p|/2}\sqrt{s_p}$, and thus

$$\begin{aligned} K_j &\lesssim \left(\sum_{p=0}^{J} 2^{-|j-p|/2}\sqrt{s_p}\right)^2 \lesssim \left(\sum_{p=0}^{J} 2^{-|j-p|/2}\right)\left(\sum_{p=0}^{J} 2^{-|j-p|/2} s_p\right) \\ &\lesssim \left(\sum_{p=0}^{J} 2^{-|j-p|/2} s_p\right) \end{aligned}$$

where in the second inequality we use Cauchy-Schwarz inequality. Therefore,

$$\begin{aligned} \Upsilon &\lesssim \max_{0\le \ell\le J} \sum_{j=0}^{J} 2^{-|j-\ell|} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p \\ &\lesssim \left(\max_{0\le \ell\le J} \sum_{j=0}^{J} 2^{-|j-\ell|}\right)\left(\max_{0\le j\le J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p\right) \\ &\lesssim \max_{0\le j\le J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p. \end{aligned}$$

■

Note that the upper bounds given in Lemmas D.2 and D.3 coincide. Therefore, we can apply Theorem 3.3 with the following upper bound for $\Gamma(S,\pi)$

$$\Gamma(S,\pi) \lesssim \max_{0\le j\le J} \frac{1}{\tilde{\pi}_j} 2^{-j} \sum_{p=0}^{J} 2^{-|j-p|/2} s_p,$$

and conclude the proof for Corollary 4.4.

D.4 Proof of Corollary 4.7

Recall that $A_0 = \phi \otimes \phi \in \mathbb{C}^{n\times n}$, where $\phi \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ is a 1D orthogonal transform. Consider a blocks dictionary made of $\sqrt{n}$ horizontal lines, i.e. for $1 \le k \le \sqrt{n}$

$$B_k = \left(\phi_{k,1}\phi, \ldots, \phi_{k,\sqrt{n}}\phi\right), \quad \text{and thus} \quad B_k^* B_k = \left(\phi_{k,i}^* \phi_{k,j} \mathrm{Id}_{\sqrt{n}}\right)_{1\le i,j\le\sqrt{n}}.$$

Now, let us fix that the signal support S is concentrated on q horizontal lines of the spatial plane. Formally,

$$S \subset \{(j-1)\sqrt{n} + \{1,\ldots,\sqrt{n}\}, j \in J\} \tag{36}$$

where $J \subset \{1,\ldots,\sqrt{n}\}$ and $|J| = q$. Therefore,

$$B_k^* B_{k,S} = \left(\delta_{j\in J}\phi_{k,i}^* \phi_{k,j} \mathrm{Id}_{\sqrt{n}}\right)_{1\le i,j\le\sqrt{n}},$$

where $\delta_{j\in J} = 1$ if $j \in J$ and 0 otherwise. In such a setting, the quantities in Definition 3.1 can be rewritten as follows:

$$\Theta(S,\pi) = \max_{1\le k\le M} \max_{1\le i\le n} \frac{\|e_i^* B_k^* B_{k,S}\|_1}{\pi_k} = \max_{1\le k\le\sqrt{n}} \max_{1\le \tilde{i}\le\sqrt{n}} \frac{|\phi_{k,\tilde{i}}| \sum_{j\in J} |\phi_{k,j}|}{\pi_k} \le \max_{1\le k\le\sqrt{n}} q \frac{\|\phi_{k,:}\|_\infty^2}{\pi_k}. \tag{37}$$

Recall that

$$\Upsilon(S,\pi) := \max_{1\le i\le n} \sup_{\|v\|_\infty\le 1} \sum_{k=1}^{M} \frac{1}{\pi_k} \left|e_i^* B_k^* B_{k,S} v\right|^2,$$

and call $(i^\star, v)$ the argument of the supremum over $\{1,\dots,n\}$ and $\{u, \|u\|_\infty \le 1\}$. Therefore,

$$\Upsilon(S,\pi) = \sum_{k=1}^{M} \frac{1}{\pi_k} \left|e_{i^\star}^* B_k^* B_{k,S} v\right|^2.$$

We can decompose $i^\star = (i_1-1)\sqrt{n} + i_2$ with i_1, i_2 integers of $\{1,\dots,\sqrt{n}\}$. We can write

$$\Upsilon(S,\pi) = \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\sum_{j=1}^{\sqrt{n}} \delta_{j\in J} \phi_{k,i_1}^* \phi_{k,j} e_{i_2}^* v[j]\right|^2 = \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} |\phi_{k,i_1}|^2 \left|\sum_{j=1}^{\sqrt{n}} \delta_{j\in J} \phi_{k,j} w_j\right|^2,$$

where $w \in \mathbb{C}^{\sqrt{n}}$ such that $w_j = e_{i_2}^* v[j]$ and $v[j] \in \mathbb{C}^{\sqrt{n}}$ is the restriction of v to the j-th horizontal line, i.e. to the components of v indexed by $\{(j-1)\sqrt{n}+1,\dots,j\sqrt{n}\}$. We can rewrite the last expression as follows

$$\begin{aligned}
\Upsilon(S,\pi) &= \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} |\phi_{k,i_1}|^2 \left|\langle e_k, \phi P_J w\rangle\right|^2 \le \max_{1\le \ell\le\sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \sum_{k=1}^{\sqrt{n}} \left|\langle e_k, \phi P_J w\rangle\right|^2 \\
&= \max_{1\le \ell\le\sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \|\phi P_J w\|_2^2 = \max_{1\le \ell\le\sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \|P_J w\|_2^2 \\
&\le \max_{1\le \ell\le\sqrt{n}} \frac{1}{\pi_\ell} |\phi_{\ell,i_1}|^2 \cdot q,
\end{aligned}$$

where in the last expression we use that $\|w\|_\infty \le 1$. Choosing ϕ as the 1D Fourier transform gives $\|\phi_{\ell,:}\|_\infty = \frac{1}{n^{1/4}}$ and choosing a uniform sampling among the $\sqrt{n}$ horizontal lines, i.e. $\pi_\ell^\star = 1/\sqrt{n}$ for $1 \le \ell \le \sqrt{n}$, leads to

$$\Gamma(S,\pi^\star) \le q,$$

which ends the proof of Corollary 4.7.

D.5 Proof of Corollary 4.9

We recall that the sampling matrix is then constructed from the full sampling matrix $A_0 \in \mathbb{C}^{n\times n}$, in the 2D setting, where $A_0 = \mathcal{F}_{2D}\Psi^*$ with $\mathcal{F}_{2D} \in \mathbb{C}^{n\times n}$ the 2D Fourier transform and $\Psi^* \in \mathbb{C}^{n\times n}$ the 2D inverse wavelet transform. Since both transforms are separable, $\mathcal{F}_{2D} = \mathcal{F}\otimes\mathcal{F}$, $\Psi = \psi\otimes\psi$, with $\otimes$ the Kronecker product and $\mathcal{F}, \psi \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ the corresponding 1D transforms. Then A_0 can also be rewritten as $A_0 = \phi\otimes\phi$, the Kronecker product of the 1D transforms $\phi := \mathcal{F}\psi^* \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$.

In this section, in order to avoid any confusion, we will denote by $\left(e_i^{(n)}\right)_{1\le i\le n}$ the canonical basis in dimension n.

In Corollary 4.9, we focus on the case where $A_0 = \phi\otimes\phi \in \mathbb{C}^{n\times n}$ is the 2D Fourier-Shannon wavelet transform, then $\phi \in \mathbb{C}^{\sqrt{n}\times\sqrt{n}}$ is the 1D Fourier-Shannon wavelets transform. Therefore, ϕ and A_0 are block-diagonal orthogonal matrices. The sensing schemes are based on horizontal lines on the 2D plane, meaning that

$$B_k = \left(\phi_{k,1}\phi \dots \phi_{k,\sqrt{n}}\phi\right),$$

for $k = 1,\dots,\sqrt{n}$. By defintion of the Fourier-Shannon transform, we have that

$$B_k^* B_k = \left(\phi_{k,\ell}^* \phi_{k,m} \mathrm{Id}_{\sqrt{n}}\right)_{1\le \ell,m\le\sqrt{n}} = \frac{1}{2^{j(k)}} \left(\delta_{\ell\in\tau_{j(k)}} \delta_{m\in\tau_{j(k)}} \mathrm{Id}_{\sqrt{n}}\right)_{1\le \ell,m\le\sqrt{n}},$$

for $k = 1, \dots, \sqrt{n}$, where $\delta_{\ell \in \tau_j} = 1$ if $\ell \in \tau_j$, and 0 otherwise.

First let us start with the evaluation of Θ. By definition of $\|\cdot\|_{\infty\to\infty}$, we have

$$\|B_k^* B_{k,S}\|_{\infty\to\infty} = \max_{1\le \ell\le n} \sup_{\substack{\|v\|_\infty \le 1 \\ v\in\mathbb{C}^n}} \left| \left(e_\ell^{(n)}\right)^* B_k^* B_k P_S v \right|.$$

Setting $\tilde{v} = \tilde{v}(k)$ the argument of the supremum in the last expression, then

$$\Theta := \max_{1\le k\le\sqrt{n}} \max_{1\le \ell\le n} \frac{1}{\pi_k} \left| \left(e_\ell^{(n)}\right)^* B_k^* B_k P_S \tilde{v} \right|,$$

Note that $\|\tilde{v}\|_\infty \le 1$. The index ℓ can be rewritten as $\ell = (\ell_1 - 1)\sqrt{n} + \ell_2$, with $1 \le \ell_1, \ell_2 \le \sqrt{n}$.

$$\begin{aligned}
\Theta &:= \max_{1\le k\le\sqrt{n}} \max_{1\le \ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \left(\phi_{k,m}\left(e_{\ell_2}^{(\sqrt{n})}\right)^*\right)_{1\le m\le\sqrt{n}} P_S\tilde{v} \right|, \\
&= \max_{1\le k\le\sqrt{n}} \max_{1\le \ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m=1}^{\sqrt{n}} \phi_{k,m}\left(e_{\ell_2}^{(\sqrt{n})}\right)^* (P_S\tilde{v})[m] \right|,
\end{aligned}$$

where $(v)[m] \in \mathbb{C}^{\sqrt{n}}$ is the restriction of the vector v to the m-th horizontal line, i.e. to the components indexed by $\{(m-1)\sqrt{n}+1, \dots, m\sqrt{n}\}$. Set $w^{|(m)} := (P_S\tilde{v})[m] \in \mathbb{C}^{\sqrt{n}}$, the restriction of $P_S\tilde{v}$ to the m-th horizontal line. Then the ℓ_2-th component of $w^{|(m)}$, written as $w_{\ell_2}^{|(m)}$ is equal to $\left(e_{\ell_2}^{(\sqrt{n})}\right)^* (P_S\tilde{v})[m]$. Note that $\left|w_{\ell_2}^{|(m)}\right| \le 1$ if $(m-1)\sqrt{n} + \ell_2 \in S$, and it is equal to 0 otherwise. Then,

$$\Theta \le \max_{1\le k\le\sqrt{n}} \max_{1\le \ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m=1}^{\sqrt{n}} \phi_{k,m} w_{\ell_2}^{|(m)} \right|. \tag{38}$$

By the properties of block-diagonality of the Fourier-Shannon transform, we have

$$\begin{aligned}
\Theta &\le \max_{1\le k\le\sqrt{n}} \max_{1\le \ell_1,\ell_2\le\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1}^* \sum_{m\in\tau_{j(k)}} \phi_{k,m} w_{\ell_2}^{|(m)} \right| \qquad (39) \\
&\le \max_{1\le k\le\sqrt{n}} \max_{1\le \ell_2\le\sqrt{n}} \frac{1}{\pi_k} \|\phi_{k,:}\|_\infty^2 \left| \sum_{m\in\tau_{j(k)}} w_{\ell_2}^{|(m)} \right| \\
&\le \max_{1\le k\le\sqrt{n}} \max_{1\le \ell_2\le\sqrt{n}} \frac{1}{\pi_k} \|\phi_{k,:}\|_\infty^2 \sum_{m\in\tau_{j(k)}} \left| w_{\ell_2}^{|(m)} \right| \\
&\lesssim \max_{1\le k\le\sqrt{n}} \frac{1}{\pi_k} \frac{1}{2^{j(k)}} s_{j(k)}^c. \qquad (40)
\end{aligned}$$

Indeed, $\sum_{m\in\tau_{j(k)}} \left|w_{\ell_2}^{|(m)}\right|$ is bounded above by $\sum_{m\in\tau_{j(k)}} \delta_{(m-1)\sqrt{n}+\ell_2\in S}$, which counts the number of intersections between S, the ℓ_2th-column and the $j(k)$ (horizontal) level, see the blue line in Figure 3. Taking the maximum over $1 \le \ell_2 \le \sqrt{n}$ leads to $\sum_{m\in\tau_{j(k)}} \delta_{(m-1)\sqrt{n}+\ell_2\in S} \le s_{j(k)}^c$.

Secondly, let us evaluate Υ. We have that

$$\Upsilon := \max_{1\le \ell\le n} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left| \left(e_\ell^{(n)}\right)^* B_k^* B_k \tilde{v} \right|^2,$$

where $\tilde{v} = \tilde{v}(\ell)$ is the argument of the supremum on the ℓ_∞ unit-ball. Using (39), with $\ell = (\ell_1 - 1)\sqrt{n} + \ell_2$, we can rewrite

$$\Upsilon = \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left| \phi^*_{k,\ell_1} \sum_{m=1}^{\sqrt{n}} \phi_{k,m} w^{|(m)}_{\ell_2} \right|^2,$$

where $w^{|(m)}_{\ell_2} := \left(e^{(\sqrt{n})}_{\ell_2}\right)^* (P_S \tilde{v})[m]$. Note again that $\left|w^{|(m)}_{\ell_2}\right| \le 1$ if $(m-1)\sqrt{n} + \ell_2 \in S$, and it is equal to 0 otherwise. By denoting $w^{|(:,\ell_2)}$ the vector with components

$$w^{|(:,\ell_2)} := \left(w^{|(1)}_{\ell_2}, \quad w^{|(2)}_{\ell_2}, \ldots, w^{|(\sqrt{n})}_{\ell_2}\right)^*, \tag{41}$$

we can rewrite the previous quantity as follows

$$\begin{aligned} \Upsilon &= \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left| \phi^*_{k,\ell_1} \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle \right|^2 \\ &= \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\phi_{k,\ell_1}\right|^2 \left| \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle \right|^2. \end{aligned} \tag{42}$$

Since ϕ is an orthogonal block-diagonal transform, we have

$$\Upsilon = \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \sum_{k \in \tau_{j(\ell_1)}} \frac{1}{\pi_k} \left|\phi_{k,\ell_1}\right|^2 \left| \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle \right|^2.$$

Choosing $\pi_k = \tilde{\pi}_j$ for $k \in \tau_j$ meaning that the probability of drawing lines is constant by levels, we can write that

$$\begin{aligned} \Upsilon &= \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \frac{1}{\tilde{\pi}_{j(\ell_1)}} \sum_{k \in \tau_{j(\ell_1)}} \left|\phi_{k,\ell_1}\right|^2 \left| \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle \right|^2, \\ &\le \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \frac{1}{\tilde{\pi}_{j(\ell_1)}} \sum_{k \in \tau_{j(\ell_1)}} \|\phi_{k,:}\|^2_\infty \left| \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle \right|^2, \\ &\lesssim \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} \sum_{k \in \tau_{j(\ell_1)}} \left| \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle \right|^2, \\ &= \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} \left\| P_{\tau_{j(\ell_1)}} \phi w^{|(:,\ell_2)} \right\|^2_2. \end{aligned}$$

Since ϕ is orthogonal and block diagonal we have $\left\| P_{\tau_{j(\ell_1)}} \phi w^{|(:,\ell_2)} \right\|^2_2 = \|P_{\tau_{j(\ell_1)}} w^{|(:,\ell_2)}\|^2_2$. Then,

$$\begin{aligned} \Upsilon &\lesssim \max_{1 \le \ell_1, \ell_2 \le \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} \left\| P_{\tau_{j(\ell_1)}} w^{|(:,\ell_2)} \right\|^2_2, \\ &\lesssim \max_{1 \le \ell_1 \le \sqrt{n}} \frac{2^{-j(\ell_1)}}{\tilde{\pi}_{j(\ell_1)}} s^c_{j(\ell_1)}, \end{aligned} \tag{43}$$

where the last step invokes that $\|P_{\tau_{j(\ell_1)}} w^{|(:,\ell_2)}\|^2_2 \le \sum_{m \in \tau_{j(\ell_1)}} \delta_{(m-1)\sqrt{n}+\ell_2 \in S} \le s^c_{j(\ell_1)}$. Note that the upper bounds (40) and (43) on Υ and Θ coincide. They lead to the following choice for $1 \le k \le \sqrt{n}$,

$$\pi_k = \tilde{\pi}_{j(k)} = \frac{s^c_{j(k)} 2^{-j(k)}}{\sum_{\ell=1}^{\sqrt{n}} s^c_{j(\ell)} 2^{-j(\ell)}} = \frac{s^c_{j(k)} 2^{-j(k)}}{\sum_{j=0}^{J} \sum_{\ell \in \tau_j} s^c_j 2^{-j}} = \frac{s^c_{j(k)} 2^{-j(k)}}{\sum_{j=0}^{J} s^c_j}.$$

Then for this particular choice, we can rewrite

$$\max(\Theta, \Upsilon) \lesssim \sum_{j=0}^{J} s_j^c.$$

To conclude, by Theorem 3.3, a lower bound on the required number of horizontal lines to acquire is thus

$$m \gtrsim \sum_{j=0}^{J} s_j^c \ln(s) \ln(n/\varepsilon).$$

D.6 Proof of Corollary 4.10

In this part, using the formalism introduced in the last section, ψ is the 1D Haar transform, and ϕ is then the Fourier-Haar's wavelet transform. In such a case, we can reuse (38) in Section D.5 to evaluate Θ:

$$\Theta \leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left| \phi_{k,\ell_1} \sum_{m=1}^{\sqrt{n}} \phi_{k,m} w_m[\ell_2] \right|.$$

Using Lemma D.1, we have for $1 \leq k, m \leq \sqrt{n}$,

$$|\phi_{k,m}| \lesssim 2^{-j(k)/2} 2^{-|j(k)-j(m)|/2}.$$

Therefore,

$$\begin{aligned}
\Theta &\leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right| \sum_{m=1}^{\sqrt{n}} \left|\phi_{k,m} w^{|(m)}_{\ell_2}\right| \\
&\leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right| \sum_{j=0}^{J} \sum_{m\in\tau_j} |\phi_{k,m}| \left|w^{|(m)}_{\ell_2}\right| \\
&\leq \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right| \sum_{j=0}^{J} \sum_{m\in\tau_j} |\phi_{k,m}| \left|w^{|(m)}_{\ell_2}\right| \\
&\lesssim \max_{1\leq k\leq\sqrt{n}} \max_{1\leq \ell_2\leq\sqrt{n}} \frac{1}{\pi_k} 2^{-j(k)} \sum_{j=0}^{J} 2^{-|j(k)-j|/2} \sum_{m\in\tau_j} \left|w^{|(m)}_{\ell_2}\right| \\
&\lesssim \max_{1\leq k\leq\sqrt{n}} \frac{1}{\pi_k} 2^{-j(k)} \sum_{j=0}^{J} 2^{-|j(k)-j|/2} s_j^c. \qquad (44)
\end{aligned}$$

Now let us study Υ. Recall the definition of $w^{|(:,\ell_2)}$ depending on ℓ_2 in (41), we can reuse (42) to have

$$\begin{aligned}
\Upsilon &= \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1} \left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2 \\
&= \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{k=1}^{\sqrt{n}} \frac{1}{\pi_k} \left|\phi^*_{k,\ell_1}\right|^2 \left|\left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2, \\
&= \max_{1\leq \ell_1,\ell_2\leq\sqrt{n}} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} \sum_{k\in\tau_j} \left|\phi^*_{k,\ell_1}\right|^2 \left|\left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2,
\end{aligned}$$

by choosing $\pi_k = \tilde{\pi}_j$ for $k \in \tau_j$, meaning that the drawing probability is constant by level. Since for $k \in \tau_j$, we have $\left|\phi^*_{k,\ell_1}\right|^2 \leq 2^{-j}2^{-|j-j(\ell_1)|}$ by Lemma D.1. Then,

$$\Upsilon = \max_{1\leq \ell_1,\ell_2 \leq \sqrt{n}} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} 2^{-j} 2^{-|j-j(\ell_1)|} \underbrace{\sum_{k\in\tau_j} \left|\left\langle \phi^*_{k,:}, w^{|(:,\ell_2)} \right\rangle\right|^2}_{=:K_j}.$$

Dealing with K_j, we can derive that

$$\begin{aligned}\sqrt{K_j} = \left\|P_{\tau_j}\phi^* w^{|(:,\ell_2)}\right\|_2 = \left\|P_{\tau_j}\phi^* \sum_{r=0}^{J} P_{\tau_r} w^{|(:,\ell_2)}\right\|_2 \quad &\leq \sum_{r=0}^{J} \left\|P_{\tau_j}\phi^* P_{\tau_r}\right\|_{2\to 2} \left\|P_{\tau_r} w^{|(:,\ell_2)}\right\|_2 \\ \lesssim \sum_{r=0}^{J} 2^{-|j-r|/2}\sqrt{s_r^c}, \quad &\end{aligned}$$

where the upper bound $\left\|P_{\tau_j}\phi^* P_{\tau_r}\right\|_{2\to 2} \lesssim 2^{-|j-r|/2}$ can be found in [AHR14a, Lemma 4.3]. Then,

$$\begin{aligned} K_k &\lesssim \left(\sum_{r=0}^{J} 2^{-|j-r|/2}\sqrt{s_r^c}\right)^2 \lesssim \left(\sum_{r=0}^{J} 2^{-|j-r|/2}\right)\left(\sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c\right) \\ &\lesssim \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c. \end{aligned}$$

Therefore,

$$\begin{aligned} \Upsilon &\lesssim \max_{1\leq \ell_1 \leq \sqrt{n}} \sum_{j=0}^{J} \frac{1}{\tilde{\pi}_j} 2^{-j} 2^{-|j-j(\ell_1)|} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \\ &\lesssim \left(\max_{1\leq \ell_1 \leq \sqrt{n}} \sum_{j=0}^{J} 2^{-|j-j(\ell_1)|}\right)\left(\max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c\right) \\ &\lesssim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c. \end{aligned} \tag{45}$$

The upper bounds (44) and (45) give

$$\max(\Theta, \Upsilon) \lesssim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c.$$

Therefore, by Theorem 3.3, a lower bound on the required number of horizontal lines is

$$m \gtrsim \max_{0\leq j\leq J} \frac{2^{-j}}{\tilde{\pi}_j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \ln(n/\varepsilon)\ln(s).$$

By choosing

$$\pi_k = \tilde{\pi}_{j(k)} = \frac{2^{-j(k)} \sum_{r=0}^{J} 2^{-|j(k)-r|/2} s_r^c}{\sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c},$$

for $1 \leq k \leq \sqrt{n}$, the lower bound on the required number of horizontal lines can be rewritten as

$$\begin{aligned} m &\gtrsim \sum_{\ell=1}^{\sqrt{n}} 2^{-j(\ell)} \sum_{r=0}^{J} 2^{-|j(\ell)-r|/2} s_r^c \cdot \ln(n/\varepsilon)\ln(s) \\ &\gtrsim \sum_{j=0}^{J} \sum_{\ell \in \tau_j} 2^{-j} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \cdot \ln(n/\varepsilon)\ln(s) \\ &\gtrsim \sum_{j=0}^{J} \sum_{r=0}^{J} 2^{-|j-r|/2} s_r^c \cdot \ln(n/\varepsilon)\ln(s) \\ &\gtrsim \sum_{j=0}^{J} \left(s_j^c + \sum_{\substack{r=0 \\ r \neq j}}^{J} 2^{-|j-r|/2} s_r^c \right) \cdot \ln(n/\varepsilon)\ln(s), \end{aligned}$$

which concludes the proof of Corollary 4.10.

References

[AHPR13] Ben Adcock, Anders C. Hansen, Clarice Poon, and Bogdan Roman. Breaking the coherence barrier: A new theory for compressed sensing. *arXiv preprint arXiv:1302.0561*, 2013.

[AHR14a] Ben Adcock, Anders C Hansen, and Bogdan Roman. A note on compressed sensing of structured sparse wavelet coefficients from subsampled fourier measurements. *arXiv preprint arXiv:1403.6541*, 2014.

[AHR14b] Ben Adcock, Anders C. Hansen, and Bogdan Roman. The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing. *Book Chapter, Compressed Sensing and its Applications, Springer (to appear), arXiv preprint arXiv:1403.6540*, 2014.

[BBW14] Jérémie Bigot, Claire Boyer, and Pierre Weiss. An analysis of blocks sampling strategies in compressed sensing. *arXiv preprint arXiv:1310.4393*, 2014.

[BH14] Alexander Bastounis and Anders C Hansen. On the absence of the rip in real-world applications of compressed sensing and the rip in levels. *arXiv preprint arXiv:1411.4449*, 2014.

[BWB14] Claire Boyer, Pierre Weiss, and Jérémie Bigot. An algorithm for variable density sampling with block-constrained acquisition. *SIAM Journal on Imaging Sciences*, 7(2):1080–1107, 2014.

[CCKW14] Nicolas Chauffert, Philippe Ciuciu, Jonas Kahn, and Pierre Weiss. Variable density sampling with continous sampling trajectories. *SIAM Journal on Imaging Sciences, in press*, 2014.

[CCW13] N. Chauffert, P. Ciuciu, and P. Weiss. Variable density compressed sensing in MRI. theoretical vs heuristic sampling strategies. In *proceedings of IEEE ISBI*, 2013.

[CP11] Emmanuel Candès and Yaniv Plan. A probabilistic and ripless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254, 2011.

[CRT06a] Emmanuel Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[CRT06b] Emmanuel Candès, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[CT06] Emmanuel Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.

[DE11] Marco F Duarte and Yonina C Eldar. Structured compressed sensing: From theory to applications. *Signal Processing, IEEE Transactions on*, 59(9):4053–4085, 2011.

[Don06] David Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[Fel08] Willliam Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008.

[FR13] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[GN08] Rémi Gribonval and Morten Nielsen. Beyond sparsity: Recovering structured representations by {\ ell}ˆ 1 minimization and greedy algorithms. *Advances in computational mathematics*, 28(1):23–41, 2008.

[Gro11] David Gross. Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566, 2011.

[HSIG13] Cédric Herzet, Charles Soussen, Jérôme Idier, and Rémi Gribonval. Exact recovery conditions for sparse representations with partial support information. *Information Theory, IEEE Transactions on*, 59(11):7509–7524, 2013.

[JKN+12] Anatoli Juditsky, Fatma Kılınç Karzan, Arkadi Nemirovski, Boris Polyak, et al. Accuracy guaranties for ℓ_1 recovery of block-sparse signals. *The Annals of Statistics*, 40(6):3077–3107, 2012.

[KW14] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Trans. Image Proc.*, 23(2):612–622, 2014.

[LDP07] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine*, 58(6):1182–1195, 2007.

[PDG15] Adam C. Polak, Marco F. Duarte, and Dennis L. Goeckel. Performance bounds for grouped incoherent measurements in compressive sensing. *To appear in IEEE Signal Processing*, 2015.

[PMG+12] Gilles Puy, Jose P. Marques, Rolf Gruetter, J. Thiran, Dimitri Van De Ville, Pierre Vandergheynst, and Yves Wiaux. Spread spectrum magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):586–598, 2012.

[PVW11] Gilles Puy, Pierre Vandergheynst, and Yves Wiaux. On variable density compressive sampling. *Signal Processing Letters, IEEE*, 18(10):595–598, 2011.

[Rau10] Holger Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[RHA14] Bogdan Roman, Anders Hansen, and Ben Adcock. On asymptotic structure in compressed sensing. *arXiv preprint arXiv:1406.4178*, 2014.

[Rom09] Justin Romberg. Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences*, 2(4):1098–1128, 2009.

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.